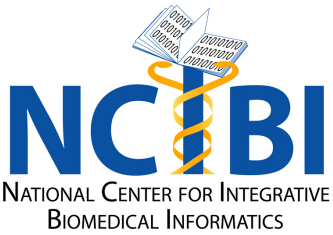
This work is supported by the
National Center for Integrative Biomedical Informatics
through NIH Grant# 1U54DA021519-01A1

Michigan Molecular Interactions (MiMI)

Users Manual

Version 3.1.0

Last updated November 25, 2008



Michigan Molecular Interactions User Manual

Table of Contents

Preface: Licensing

Terms of Use

MiMI is both a web service that integrates data and the application of research and open source software to the problem of supporting that web service. For use of the supporting software, please refer to that list (see Software Products Used below). The data provided on this website has been compiled and merged from multiple sources. See Sources of Data.

The data integration service is supplied under the conditions of the original data sources and the specific terms of use for MiMI. Access to this website is provided free of charge. Permission is granted to use this software and data internally only, so long as no fee is charged, usage of this website is cited in any resulting publications involving results from such use, and so long as the name of the University of Michigan is not used in any advertising or publicity pertaining to such use without specific, written prior authorization. Permission to redistribute this data in any form is specifically not granted.

The Regents of the University of Michigan does not check this data for errors or omissions, and by its nature, the data included herein likely contains errors and omissions. Access and use is provided as is, without representation as to its fitness for any purpose, and without warranty of any kind, either express or implied, including without limitation the implied warranties of merchantability and fitness for a particular purpose. The Regents of the University of Michigan shall not be liable for any damages, including special, indirect, incidental, or consequential damages, with respect to any claim arising out of, or in connection with, the use of this website or data, even if it has been or is hereafter advised of the possibility of such damages.

Software Products Used

MiMI and the MiMI Web Site are built on a number of standard, open source software tools.

(1) Timber (http://www.eecs.umich.edu/db/timber/). In the TIMBER project we are exploring the issues involved in storing XML in native format. We recognize XML documents to be trees and built a system to manipulate collections of trees. In doing so, we attempt to avoid the pitfall of "instance-at-a-time" navigational access. Rather, we attempt to bring to bear the core ideas of database technology, such as declarative querying, a bulk algebra, and cost-based query o ptimization.

(2) Tomcat (http://tomcat.apache.org/). Apache Tomcat is the servlet container that is used in the official Reference Implementation for the Java Servlet and JavaServer Pages technologies. The Java Servlet and JavaServer Pages specifications are developed by Sun under the Java Community Process.

(3) Lucene (http://lucene.apache.org/java/docs/). Apache Lucene is a high-performance, full-featured text search engine library written entirely in Java. It is a technology suitable for nearly any application that requires full-text search, especially cross-platform.

(4) Cytoscape (http://www.cytoscape.org/). Cytoscape is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data.

(5) ClairLib: The Clair library is a suite of open-source Perl modules intended to simplify a number of generic tasks in natural language processing (NLP), information retrieval (IR), and network analysis (NA). Its architecture also allows for external software to be plugged in with very little effort.

(6) GIN: GIN (Gene Interaction Network) is a system for browsing articles and molecule interaction information. What makes GIN stand out from other similar systems is that it uses automated methods (such as dependency parsing) to mine the text for relevant information (such as protein interactions) and computes statistics for the interaction network.(7) Gene2MeSH: Gene2MeSH is an automated annotation tool that associates Medical Subject Heading (MeSH) terms with genes using the National Library of Medicine's PubMed literature database. The significance of association between genes and MeSH terms is evaluated using Fisher.s exact test and displayed in an interface in order of significance score. Users may search by gene name or MeSH term and view or download results via the web interface.

(8) SAGA/TALE: SAGA stands for Substructure Index-based Approximate Graph Alignment. It is an efficient tool for approximate subgraph matching. SAGA allows users to match a query graph against a large database of graphs. At the core of SAGA is a flexible graph distance model that incorporates node approximate matching as well as approximate structure matching. A powerful indexing method is implemented to speed up the matching process. TALE is the next generation of SAGA, designed to perform efficient alignment between pairs of large graphs.

(9) Shannon NLP Tools/PMCOA: A relational database that integrates data from PubMed, NCBI's Gene database, and natural language processing (NLP) with the full PMCOA database.

How to Cite MiMI

Please use the following citations for this web site:

1. Jayapandian, M, Chapman, A, Tarcea, VG, Yu, C, Elkiss,A, Ianni, A, Liu,B, Nandi, A, Santos, C, Andrews, P, Athey, B, States, D,. Jagadish, HV: Michigan Molecular Interactions (MiMI): Putting the Jigsaw Puzzle Together. **Nucleic Acids Research**, 2007, Vol. 35, Database issue D566-D571.

   The URL: http://nar.oxfordjournals.org/cgi/content/full/35/suppl_1/D566

   PMID: 17130145

   The abstract: *Protein interaction data exists in a number of repositories. Each repository has its own data format, molecule identifier and supplementary information. Michigan Molecular Interactions (MiMI) assists scientists searching through this overwhelming amount of protein interaction data. MiMI gathers data from well-known protein interaction databases and deep-merges the information. Utilizing an identity function, molecules that may have different identifiers but represent the same real-world object are merged. Thus, MiMI allows the users to retrieve information from many different databases at once, highlighting complementary and contradictory information. To help scientists judge the usefulness of a piece of data, MiMI tracks the provenance of all data. Finally, a simple yet powerful user interface aids users in their queries, and frees them from the onerous task of knowing the data format or learning a query language. MiMI allows scientists to query all data, whether corroborative or contradictory, and specify which sources to utilize. MiMI is part of the National Center for Integrative Biomedical Informatics (NCIBI) and is publicly available at:* http://mimi.ncibi.org/.

2. Tarcea VG, Weymouth T, Ade A, Bookvich A, Gao J, Mahavisno V, Wright Z, Chapman A, Jayapandian M, Ozgür A, Tian Y, Cavalcoli J, Mirel B, Patel J, Radev D, Athey B, States D, Jagadish HV; Michigan molecular interactions r2: from interacting proteins to pathways; Nucleic Acids Res. 2008 Oct 31. [Epub ahead of print]

   The URL: http://nar.oxfordjournals.org/cgi/content/full/gkn722v1

   PMID: 18978014

   The abstract: *Molecular interaction data exists in a number of repositories, each with its own data format, molecule identifier and information coverage. Michigan molecular interactions (MiMI) assists scientists searching through this profusion of molecular interaction data. The original release of MiMI gathered data from well-known protein interaction databases, and deep merged this information while keeping track of provenance. Based on the feedback received from users, MiMI has been completely redesigned. This article describes the resulting MiMI Release 2 (MiMIr2). New functionality includes extension from proteins to genes and to pathways; identification of highlighted sentences in source publications; seamless two-way linkage with Cytoscape; query facilities based on MeSH/GO terms and other concepts; approximate graph matching to find relevant pathways; support for querying in bulk; and a user focus-group driven interface design. MiMI is part of the NIH's; National Center for Integrative Biomedical Informatics (NCIBI) and is publicly available at:* http://mimi.ncibi.org.

Source	Terms of Use
BIND	There are no license conditions attached to the use of BIND if you are using BIND data for internal research purposes. Unleashed Informatics Limited holds an exclusive commercial license to intellectual property including U.S. patent number 6,745,204 - "System for electronically managing, finding, and/or displaying biomolecular interactions" - also known as the BIND patent. The USPTO document can be found here. (http://patft.uspto.gov/netahtml/PTO/srchnum.htm) For-profit organizations selling a biomolecular interaction software system or employing such a software system specification in a product for sale which falls under the claims of the above patent will require a commercial, fee-based license from Unleashed Informatics. Academic and commercial users of BOND will be unaffected by the enforcement of this patent.
BioGRID	http://www.thebiogrid.org/viewdocument.php?documentid=6
CCSB at Harvard	Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M.; Towards a proteome-scale map of the human protein-protein interaction network; Nature. 2005 Oct 20;437(7062):1173-8. Epub 2005 Sep 28. PMID: 16189514
Califano Lab	Email mimi-help@umich.edu for information
cPath	http://cbio.mskcc.org/software/cpath/
DIP	http://dip.doe-mbi.ucla.edu/termsofuse.html
GIN	http://gin.ncibi.org:8080/gin/about.jsp
GO	http://www.geneontology.org/GO.cite.shtml
HPRD	http://www.hprd.org
IntAct	http://www.ebi.ac.uk/intact/site/index.jsf
InterPro	http://www.ebi.ac.uk/interpro/User-FAQ-InterPro.html
IPI	http://www.ebi.ac.uk/IPI/IPIhelp.html
KEGG	http://www.genome.jp/kegg/kegg1.html
Max Delbreuck Center	Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksöz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE.; A human protein-protein interaction network: a resource for annotating the proteome.; Cell 2005 Sep 23;122(6):957-68. PMID: 16169070
MiBLAST	http://www.eecs.umich.edu/miblast/download.html
NCBI Gene	http://www.ncbi.nlm.nih.gov/About/disclaimer.html
Organelle DB	http://organelledb.lsi.umich.edu/index.php
OrthoMCL DB	http://www.orthomcl.org/cgi-bin/OrthoMclWeb.cgi?rm=orthomcl#Acknowledgement
PFam	http://pfam.sanger.ac.uk/help?tab=helpReferencesBlock
PMCOA	NCIBI has parsed PubMed and POubMed Central literature according to several existing and some newly developed algorithms. This BioNLP database of parsed sentences, Genes, MeSH headings and relational parts of speech is a foundation database and accessible to all our current tools.
ProtoNet	http://www.protonet.cs.huji.ac.il/prototeam.php?global=protonet|no|5|51|lifetie|1|2|2
PubMed	See NCBI Gene
PubMed NLP Mining	See MiMI Terms of Use
Reactome	http://www.reactome.org/

1.1 What is MiMI?

MiMI (Michigan Molecular Interactions) is a tool provided by the National Institutes of Health's National Center for Integrative Biomedical Informatics (NCIBI).

MiMI provides access to the knowledge and data merged and integrated from numerous protein interactions databases. It augments this information from many other biological sources. You can link out to these other databases and auxiliary sources from MiMI, as well. Please see Sources of Data for the complete list.

MiMI merges data from these sources with "deep integration" (see The MiMI Merge Process section) into its single database. A simple yet powerful user interface enables you to query the database, freeing you from the onerous task of having to know the data format or having to learn a query language.

MiMI displays results of your queries in easy-to-browse interfaces and provides you with workspaces to explore and analyze the results. Among these workspaces is an interactive network of protein-protein interactions displayed in Cytoscape and accessed through MiMI via a MiMI Cytoscape plug-in.

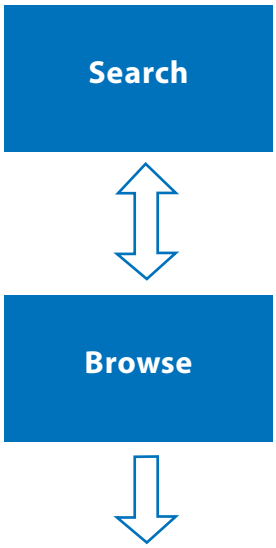
In addition, MiMI provides you with PubMed IDs of articles that support entries in the database and, in many cases, with specific sentences in those articles with gene names highlighted. This makes it easy for you to see the relevant snippet of the original source without having to access the actual article and read it.MiMI also includes information about pathways in which gene products participate. You can easily move from gene to pathway to genes in the pathway, as needed.

1.2 What Tasks Can You Perform?

With MiMI you can explore publicly available data on genes and gene products and find relationships based on biological concepts, canonical pathways, and semantic text mining. From this rich array of data, MiMI helps you uncover previously unknown knowledge within and across organisms. Insights you gain can lead to novel hypotheses about mechanisms of diseases or other biological processes that you can test through further experimentation.

MiMI merges data from these sources with "deep integration" (see How MiMI Merges Data section) into its single database. A simple yet powerful user interface enables you to query the database, freeing you from the onerous task of having to know the data format or having to learn a query language. MiMI allows you to query all data, whether corroborative or contradictory, and specify which sources to utilize.

MiMI displays results of your queries in easy-to-browse interfaces and provides you with workspaces to explore and analyze the results. Among these workspaces is an interactive network of protein-protein interactions displayed in Cytoscape and accessed through MiMI via a MiMI Cytoscape plug-in.



On the search pages, look for single genes, keywords, lists of genes or specific interactions

Construct search phrases with AND, OR, NOT

Quickly retrieve the records for all genes that match your search traits in one or more of the
fields that MiMI stores describing genes (e.g. name, description, GO annotations, and so on.)

See a summary of each gene retrieved

Move to the gene's details by clicking on the gene name.

Confirm from details if what you expect to see is displayed

Refine queries by changing search phrases

Analyze muti-variate relationships

Find/validate relationships of interest

On the details page for a selected gene you can: explore its interactions, the pathways in which it resides, associated compounds related to metabolic reactions, and literature about the gene and its interactions.

Link out as needed for more data to validate and uncover information about previously unknown relationships or attributes

Locate, skim and download potentially relevant articles

Characterize the gene through enriched MeSH terms in Gene2MeSH.

Scan NLP-extracted passages relevant to formulating a hypothesis. Extend this search and analysis of relevant passages by using GIN and Cytoscape.

Quickly envision a network of a selected gene's interactors by opening Netbrowser.

Infer causes contextually and visually

Move to Cytoscape and carry over all the MiMI data through the MiMI plug to conduct in-depth exploratory analysis of interactions and their biological meaning.

Use Cytoscape's functionality and MiMI plug in's features to find sub-graphs of interest and to match them to pathways from KEGG

Find NLP-extracted passages on interactions by right-clicking edges

Integrate files on expression data with the interactions for analysis.

Navigate across sources and tools

Save and download

Navigate to outside biological knowledge bases through linkouts.

Navigate to increasing detail within MiMI through internal links

Use integrated tools in the MIMI suite to: enhance your search on literature, find NLP mined passages, MeSH traits, and other genes associated with your gene in the literature, and locate interactive networks of protein-protein interactions.

Save the tables of details MiMI displays in csv, Excel, or XML formats

Download in PSI-MI format

1.3 What Attributes Can You Analyze?

MiMI provides the following data. See more detail in the sections on Information Found on the screen.

Genes	Interactions
Biological Processes	Gene Names
Cellular Components	Provenance Sources
Chromosome	Experiment
Description	Interaction Information
Gene Name	Literature
Gene ID	
Literature on Gene	
Organism	
Other Names	
Map Locus	
Molecular Functions	
Type	

1.4 What is Unique About MiMI?

MiMI gives you numerous ways to access and analyze more information than you can get from any one protein interaction source. For instance:

- MiMI has vetted data on genes, attributes, interactions, literature citations, annotated text extracts through natural language processing (NLP), and pathways.
- Unlike in individual resources, you can use many different synonyms to find a protein and any number of protein identifiers. You can query all fields for your search term or only specified fields (e.g. searching for p53 only in the Gene Name fields).
- You can integrate analysis with other NCIBI tools to examine overrepresented MeSH terms for genes of interest, to read additional NLP-mined text passages, and to explore interactive graphics of networks of interactions.
- You can link out to PubMed or NCIBI's MiSearch interface to PubMed for better relevance rankings.
- You can quickly see the missing information across databases from search result displays.
- You can track displayed data to its source (provenance) to help you gain confidence in the details and to identify contradictory information in the sources and make judgments about it.

2.0 The MiMI Merge Process

Protein interaction data exists in a number of repositories. Each repository has its own data format, molecule identifier, and supplementary information. MiMI assists scientists searching through this overwhelming amount of protein interaction data. MiMI gathers data from well-known protein interaction databases and deep-merges the information.

Utilizing an identity function, molecules that may have different identifiers but represent the same real-world object are merged. Thus, MiMI allows the user to retrieve information from many different databases at once, highlighting complementary and contradictory information.

There are several steps needed to create the final MiMI dataset. They are:

Because this is an automated process, and no curation occurs, any errors or misnomers in the original data sources will also exist in MiMI. For example, if a source indicates that the organism is unknown, MiMI will as well.

If you find that a molecule has been incorrectly merged under a gene record, please contact us immediately. Because MiMI is completely automatically generated, and there is no data curation, it is possible that we have merged molecules with gene records incorrectly. If made aware of the error, we can and will correct the situation. Please report any problems of this kind to mimi-help@umich.edu.

2.1 Rules and Assumptions

MiMI is not merged by 'experts' - everything is done automatically. When you look into genes of interest it is important to understand why your target gene contains the attributes and associated annotations; you also may want to know why conflicting data are displayed in the merged data. Some of the assumptions and rules that MiMI uses in deep merging that may be relevant to understanding and being confident in its displays include the following:

Source differences in quality

Not all sources are created equal. While each source has a particular strength, each also has its drawback. Some important things to know about the curation processes in these different databases that might affect your interpretations and confidence include:

- BIND - Has a large number of interactions, many based on high-throughput data.
- IntAct - Lists molecules as 'interacting' when they co-locate.

Source content

Not only do sources have different strengths and weaknesses, they also organize content differently. For example, IntAct will associate publications with individual molecules while BIND associates publications with interactions only. MiMI remains true to individual sources and how they assign publications to molecules or interactions.

Source data not used

In general, MiMI uses all data from each source. However, there are some data from each source that are not incorporated into MiMI. For instance, sequence data is never stored within the MiMI dataset.

Human Curation

Some protein interaction data sources are created by human curators. These curators read relevant publications and manually enter information into protein interaction databases. These curators provide an invaluable service by filtering information and reducing large papers into a few distinct, salient facts. Unfortunately, these curators are also human and as such, they may make human assumptions and decisions. For example, human curators may gravitate towards a favorite protein name instead of listing all known synonyms. Additionally, each source dataset has a subset of identifiers used: For example DIP concentrates on using SwissProt identifiers; while BIND uses GI.

2.2 Sources of the Displayed Data

The literature records in MiMI come from the NCIBI PubMed NLP database, a relational database integrating data from the Unified Medical Language System (UMLS), Medical Subject Headings (MeSH), NCBI's Gene database, and natural language processing (NLP) with the full NLM PubMed data set as distributed to licensees. NLP techniques are used to split PubMed abstracts into sentences, identify gene/protein names within those sentences, and parse the sentences to identify word parts-of-speech, phrase structure, and grammatical relationships. The NCIBI PubMed NLP database is updated nightly.

The literature for GIN comes from the NCIBI PMCOA (Pubmed Central Open Access) database, a relational database that integrates data from PubMed, NCBI's Gene database, and natural language processing (NLP) with the full PMCOA database.

Literature on protein interactions comes from querying data from the NCIBI PubMed NLP database for gene tags that are co-located within the same sentence.

2.3 Provenance

To help scientists judge the usefulness of a piece of data, MiMI tracks the provenance of all data and gives users a link to the sources. By presenting the sources of data, MiMI builds a user's own expertise and lets you judge the data based on what they know about the source. The NCIBI MiMI database updates from its sources monthly.

3.0 Selecting and Using a Search Method

MiMI allows you to search using several different methods, each detailed in subsections below. You access each method by clicking on the tab named for the method at the top of the screen.

Free Text	One or more keywords, gene symbols or NCBI Entrez gene IDs. Recommended for newcomers. The MiMI Free Text search uses the open source Lucene search engine for efficient full-text search capabilities. If your search term appears in any of the gene fields, the gene record will be retrieved and displayed in the results.
Gene List	Two or more gene names or NCBI Gene IDs typed into the search box or imported as a file. If your search terms occur in the field you specify - gene name or ID, the gene record will be retrieved and displayed.
Query Interactions	A gene name or ID typed into each individual search box.

3.1 Free Text Search

What is free text search?

Using the **Free Text Search** field at the top of the main search page, you can enter a single keyword, gene symbol, or Entrez Gene ID and retrieve matching genes. This version of MiMI does not warn you ahead of time if the term is "acceptable" according to Lucene rules. For problems, MiMI responds: "Processing error".

To retrieve results, MiMI searches all fields within the MiMI dataset that relate to genes. This version of MiMI does not yet search and retrieve results on Interactions (e.g. Interaction Type).

☛ **TIP:** To search for a keyword, gene symbol or gene ID that occurs only in one field, query by a specified field name, e.g.: **cellularComponent:nucleus**

Gene-related fields that MiMI searches include the following:

Field Name	Content
biologicalProcess	GO annotations
cellularComponent	GO annotations
externalref	External Reference
gdesc	Gene descriptions from free text mining and provenance sources.
geneid	From NCBI Entrez Gene
genesymbol	Gene name from Entrez Gene
genetype	Distinctions such as protein, DNA
Interactioncount	The number of interactions associated with a gene's products
moldescription	Molecule descriptions from free text mining and provenance sources
molecularFunction	GO annotations for Function
molname	Molecular Name
othernames	Gene synonyms, aliases
organism pubcount	Organism
pubcount	The number of documents related to a given gene or pairwise interaction
taxid	The numeric ID of an organism

Table 3.1. Gene related fields on which MiMI searches to retrieve results from free text searches

Entering multiple terms: search logics and filters

Free text search lets you do the following:

- Enter multiple terms
- Specify search logic with Boolean operators, nesting parentheses, or quotation marks
- Narrow searches to just one field by specifying a field name for the given term
- Run a wildcard search (using *) using only part of the spelling of a term
- Filter through a pull down list to any one organism type

❖ **Constraints:**

- Boolean operators must be capitalized
- Search terms are not case-sensitive
- The free text search treats commas as an OR
- The pull-down Organism filter permits only one choice. You may either search on All organisms or on any one but you may not search on a specified subset of organisms.

If desired search on multiple terms using Lucene logic and get the following results. For more on Lucene queries, see: http://lucene.apache.org/java/2_4_0/queryparsersyntax.html

You enter	You get:
AND *e.g. insulin AND receptor*	Records that contain all of the terms in any of the gene-related fields (intersecting). Results may be sparse or null.
OR *e.g insulin OR receptor*	Record that contain any of the terms in any of the gene-related fields (union). Results may be numerous
NOT *e.g.csf AND NOT csf3*	Records that contain specified terms except those specified after the NOT
Phrases Multiple terms unconnected by AND/OR *e.g insulin receptor*	Records in which any of the terms in the phrase or the phrase itself occur in any gene-related field. *e.g. records that reference insulin, receptor, insulin receptor*
Quotation marks Multiple terms enclosed in quotations e.g. *"insulin receptor"*	Records that have only the exact phrase in any of the gene related fields
Mixed AND and OR Multiple terms connected by AND's and OR's *e.g. Insulin OR Receptor AND Oxidation*	Records that satisfy – reading left to right – the combinations specified. e.g. Records that have either insulin or receptor in a gene-related field along with oxidation..
Field name statements Multiple terms with one or more specified by a MiMI gene-related field name e.g. *"insulin receptor" AND cellularComponent:nucleus* *e.g. geneid:1436.*	In the first example: Records for genes localized in the nucleus (with no attention to nucleus occurring in any other fields) combined with records that contain the phrase insulin receptor in any of the fields: NOTE: Searching only on the Organism field can be done through the pull down list instead of a query.
Complex search statements Multiple terms in compound and nested relationships.: *e.g. "insulin receptor" AND cellularComponent:nucleus AND (organism:homo* OR organism:dro)**	In the example: Records for genes localized in the nucleus for either humans or fruit flies that also specifically reference insulin receptor.
Wildcards Use after prefixes, e.g. phosphor*, WEE* Use in the middle of a term, e.g. c?f1r	Records that have any combinations that match your root entry.
Commas Multiple keywords, gene IDs or symbols, separated by commas, no spaces	Records that contain all of the terms in any of the gene-related fields. MiMI treats commas as an OR in free text searching

Step by step procedures

1. Enter your search term or terms in the search bar – a keyword, gene symbol, gene ID, or any combination of the three. (Figure 3.1)
2. Select an organism from the pull down list or choose All organisms. You can choose only one organism in this version of MiMI. The default is Homo Sapiens.

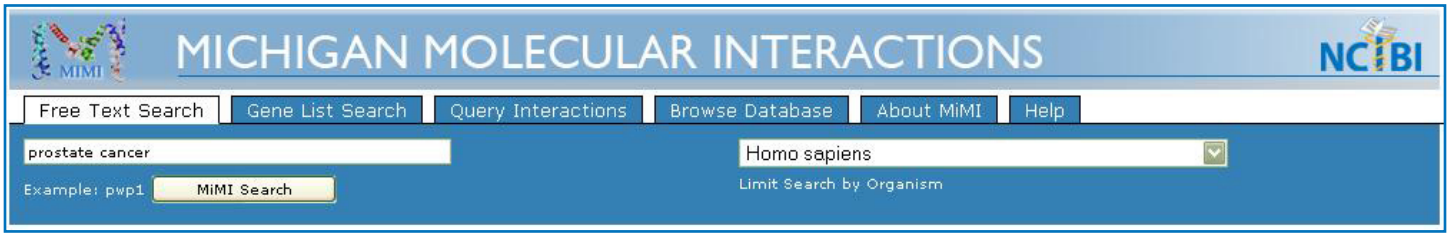


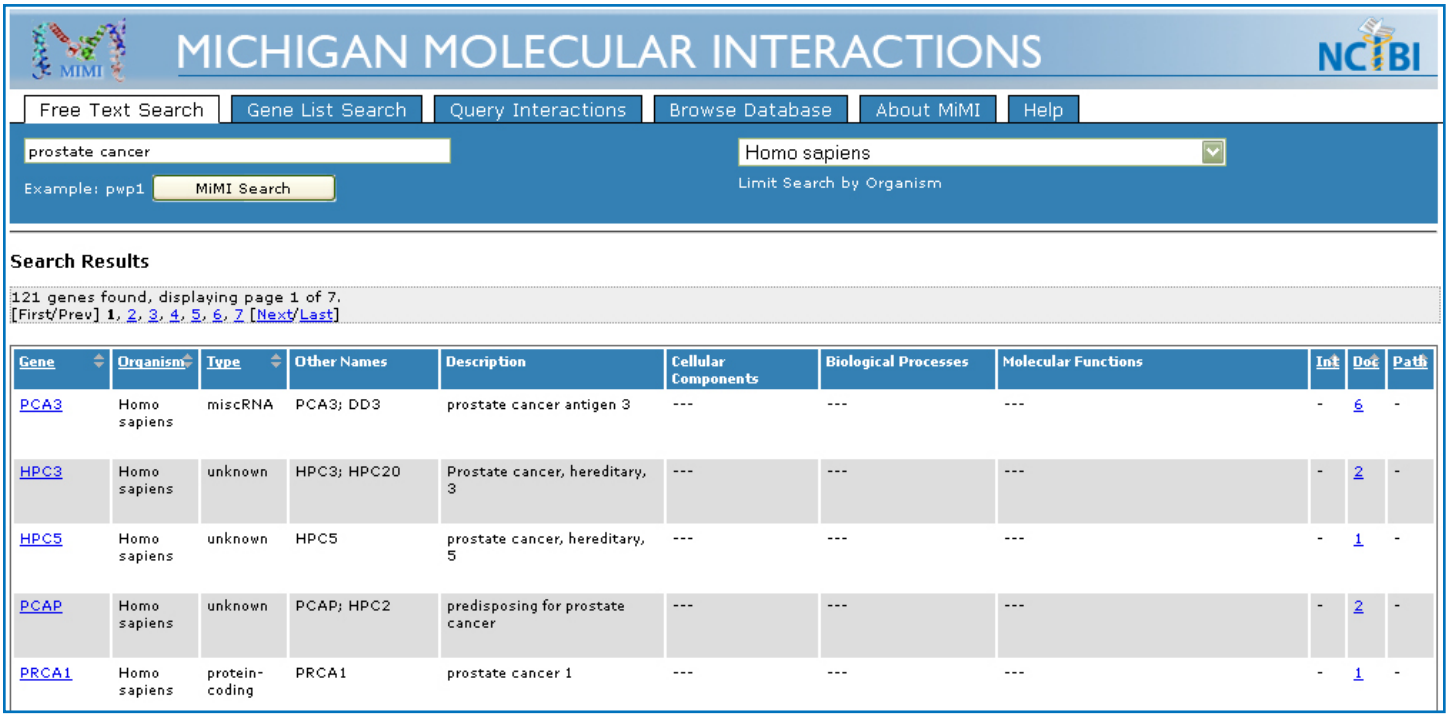
Figure 3.1 Search box and Organism Filter.

3. Click on the MiMI Search button. The search results appear (Figure 3.2), giving a summary of each gene record that fits your search criteria. For example, searching on terms prostate cancer retrieves all gene records in which the words prostate or cancer appear in one of the fields associated with the gene. Figure 3.2 shows the first page of gene records retrieved.

   From this summary, you can see gene names, gene aliases, descriptions, and GO annotations. You also can see the number of interactors the gene and its molecular products have, the number of articles referencing the gene, and the number of pathways containing it.

   If no record meets your search criteria or if the search does not meet Lucene rules, MiMI provides a message.

4. Click on the gene name (hyperlinked) to view more details for that gene.



Gene	Organism	Type	Other Names	Description	Cellular Components	Biological Processes	Molecular Functions	Int	Doc	Path
PCA3	Homo sapiens	miscRNA	PCA3; DD3	prostate cancer antigen 3	---	---	---	-	6	-
HPC3	Homo sapiens	unknown	HPC3; HPC20	Prostate cancer, hereditary, 3	---	---	---	-	2	-
HPC5	Homo sapiens	unknown	HPC5	prostate cancer, hereditary, 5	---	---	---	-	1	-
PCAP	Homo sapiens	unknown	PCAP; HPC2	predisposing for prostate cancer	---	---	---	-	2	-
PRCA1	Homo sapiens	protein-coding	PRCA1	prostate cancer 1	---	---	---	-	1	-

Figure 3.2. Search Results Screen summarizing retrieved genes.

3.2 Gene List Search

What is gene list search?

You can search by gene list in one of two ways:

- Type (or copy-paste) a linear list of gene symbols or IDs separated by spaces or returns into the text box. If you enter only one gene symbol or ID, MiMI will still conduct and complete your search.
- Upload a .txt file listing your genes symbols or IDs, with 1 entry per line.

When MiMI searches for your list of genes in its dataset, it returns all gene records that contain any one of your listed genes in the Gene Name or Gene ID field. You can use wildcards with prefixes (e.g. pw*) or in the middle (14?6) of gene names or IDs.

❖ Constraints

- Genes must be listed linearly (separated by spaces), not strings separated by commas.
- MiMI searches only the Gene Name or Gene ID field, depending on which you enter. It does not search the Other Gene Name field or any other field, including Aliases.
- MiMI treats the list as ORs, providing all records that have a match to any of the names or IDs. (See Figure 4.1)
- You can filter by only one Organism through the Organism pull down list.

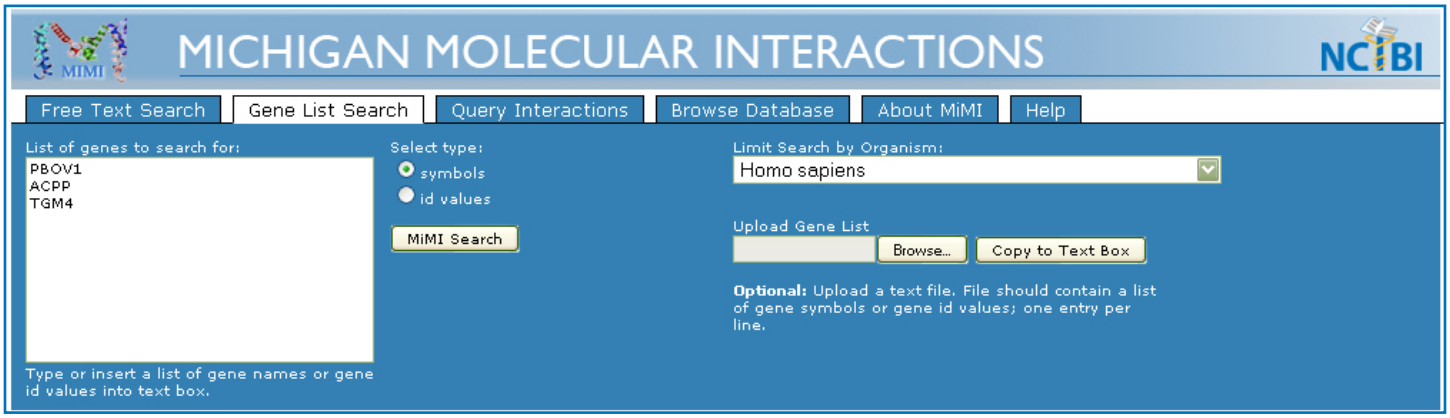


Figure3.3 Gene List Search entry form.

Step by step procedures

1. Click the Gene List Search tab. The display in Figure 3.3 appears.
2. Enter a list in the text box: Type or paste in a linear list of gene symbols or IDs. Click the appropriate radio button next to the box: Either: Symbols or ID values

   Or upload a list: Use the Upload Gene List section.

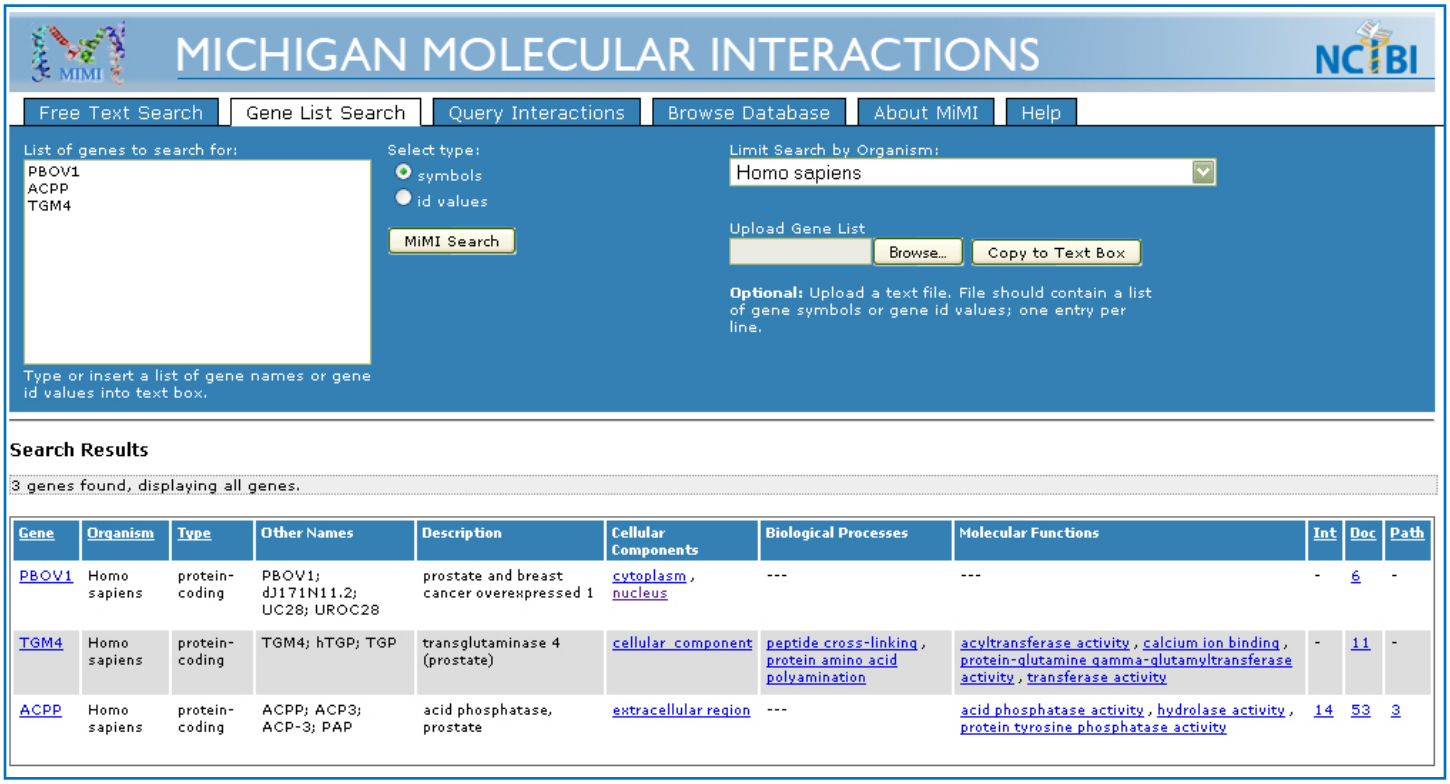
   Browse to the text file with the list of gene symbols or gene ID values.

   Select the file. Its path appears in the Browse window.

   Click the Copy to Text Box button. The genes from your file are now displayed in the box. Click the appropriate radio button next to the box: Either: Symbols or ID values.
3. Filter by an Organism from the pull down list or choose "All organisms." You can choose only one organism in this version of MiMI. The default is Homo Sapiens.
4. Click the MiMI Search button to submit your search. The Search Results screen appears. Records of all genes matching the gene names (or IDs) will appear.

   From this summary, you can see gene names, gene aliases, descriptions, and GO annotations. You also can see the number of interactors the gene and its products have, the number of articles referencing the gene, and the number of pathways containing it.

   If no records meet your search criteria or if the search does not meet Lucene rules, MiMI provides a message.
5. Click on a gene name (hyperlinked) to view more details for that gene.



Gene	Organism	Type	Other Names	Description	Cellular Components	Biological Processes	Molecular Functions	Int	Doc	Path
PBOV1	Homo sapiens	protein-coding	PBOV1; dJ171N11.2; UC28; UROC28	prostate and breast cancer overexpressed 1	cytoplasm, nucleus	---	---	-	6	-
TGM4	Homo sapiens	protein-coding	TGM4; hTGP; TGP	transglutaminase 4 (prostate)	cellular_component	peptide cross-linking, protein amino acid polyamination	acyltransferase activity, calcium ion binding, protein-glutamine gamma-glutamyltransferase activity, transferase activity	-	11	-
ACPP	Homo sapiens	protein-coding	ACPP; ACP3; ACP-3; PAP	acid phosphatase, prostate	extracellular region	---	acid phosphatase activity, hydrolase activity, protein tyrosine phosphatase activity	14	53	3

Figure 3.4 Search Results screen summarizing retrieved genes.

3.3 Query Interactions Search

What is query interaction search?

By typing in the gene symbols of two genes, you can quickly see results in the form of detailed information about the interaction - if MiMI records, in fact, show that the two genes interact. If no interactions are recorded, MIMI gives a message that nothing is found to display. MiMI defines something as an interaction when genes or gene products have been cited in the merged data sources to be interacting and when NLP-based text mining shows the pair occurs in the literature connected by "interaction-type words/verbs" (e.g. binds to, regulates).

Retrieved records may represent many types of interaction-relationships involving the pair you specify in your search, e.g. physical binding, regulatory, protein complexes.

Use this search method if you want results to show detailed information about all the interactors associated with any one gene. If you leave the second interactor text box blank, MiMi treats it as a wildcard and finds and displays all interactors with the gene products for the gene symbol or ID you entered.

❖ **Constraints:**

- You must enter a gene symbol. Any other entry results in an error/null result set.
- MiMI only searches the Gene Name field for the entries you submit. This version does not search the "Other Gene Name" field so it is important to type in the gene symbol that MiMI recognizes as the main name.
- No wildcard entries are permitted. Unlike other search modes, this one does not use Lucene.

You can filter by Organism, Interaction type, or both by using the pull down lists to the right of the text boxes. This version of MiMI allows only one filter term at a time in each of these fields. You cannot filter on any other fields.

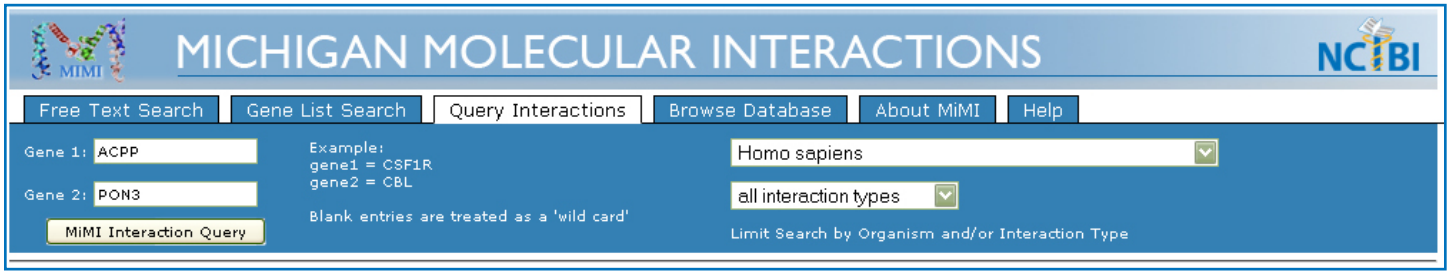


Figure 3.5 Query Interaction Search entry form

Step by step procedures

1. Click the Query Interaction Search tab. The display in Figure 3.5 appears.
2. Enter a specific pairwise interaction: Type a gene symbol into the Gene 1 and Gene 2 boxes.

   To retrieve all interactors with Gene 1, type a gene symbol into the Gene 1 box. Leave the Gene 2 box blank.
3. Filter by an Organism from the pull down list or choose "All organisms." You can choose only one organism in this version of MiMI. The default is Homo Sapiens.
4. Filter by interaction type from the pull down list or choose "All Interaction Types." The default is "All Interaction Types."

   The Interaction Types selector let you specify types of experiments finding the interaction (e.g. yeast two hybrid indicating physical binding but a high false positive rate) as well as such defining traits as a reaction, a complex, or protein-RNA relationship. Interaction types are explained in more detail in Section 5.3.
5. Click on the MiMI Interaction Query button to submit your search. A summary of interactions appears. (See Figure 3.6)

   Details include, when available, GO annotations and Interaction Type

   If no records meet your search criteria or if the search does not meet Lucene rules, MiMI provides a message.
6. Click on View (hyperlinked) to see more details for any of the interactions.

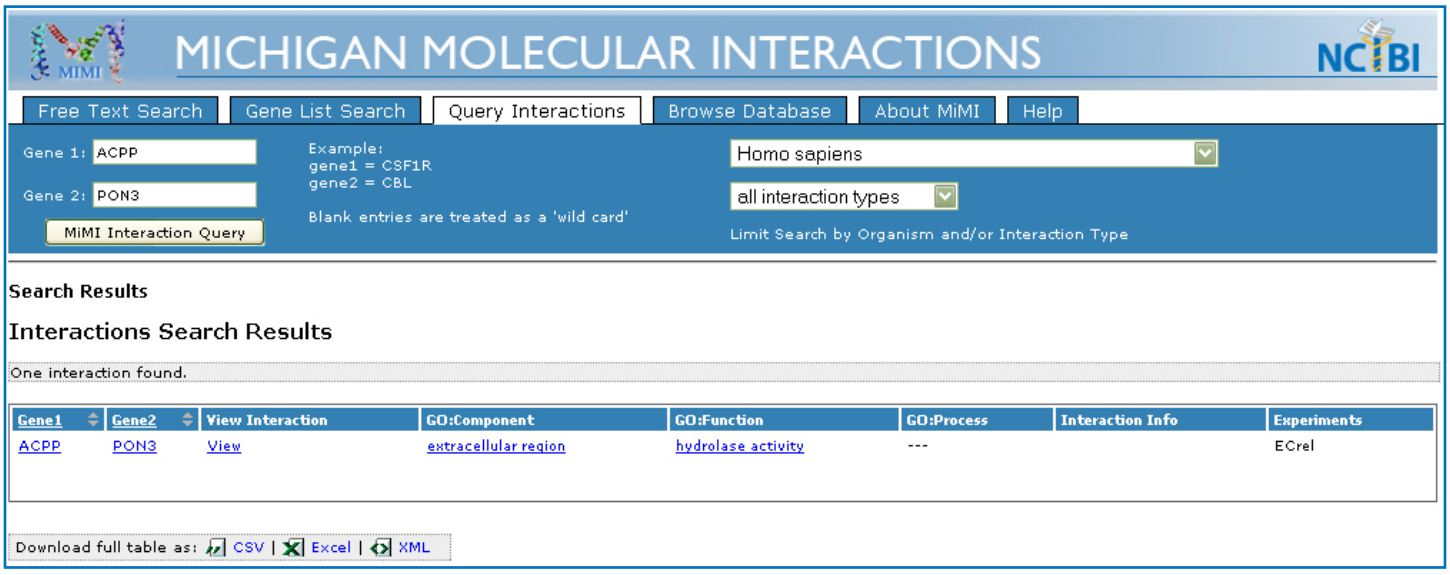


Figure 3.6 Search Results screen summarizing retrieved interactions.

3.4 Strategies for Changing Results After Searching

If no search result appears after your query or if displayed results do not contain an entry of interest to you, reformulating your search may be of use. Understanding how MiMI merges data may also be helpful in reformulating your search strategy or explaining your search results. Please refer to the The MiMI Merge Process section.

Relaxing constraints

Your query may be too specific. If you searched for: *"insulin receptor" AND cellularComponent:nucleus AND molecularFunction:transcription AND (organism:homo* OR organism:dro*)* then your search may be too specific. Try relaxing some constraints. Remember, MiMI does not populate every field. If you are only interested in a protein that occurs in the nucleus, but MiMI does not contain that information, a search that is restricts the protein to the nucleus may not produce results. Drop search terms until you have a good result set.

Dealing with Unknowns

If your search does not contain the entries you sought, consider broadening your search by eliminating unknowns. For instance, some proteins have dozens of names. However, MiMI may not be aware of the exact name you specified. Try combining a base name and using a wildcard to fill in the rest. For example, if you were looking for HNF4-alpha and typed: *hnf4_alpha*, no results are returned. However, by typing *hnf4**, several versions of HNF4-alpha are returned.

4.0 Interacting with the Search Results Screen

4.1 Step by step procedures

You can interact with the search results screen in numerous ways:

1. Go to more details: Click hyperlinks in the following columns to move to more data when you search by free text or gene list methods.

   Gene name → Click a name and see extensive details on the Gene Details page

   Int → Click a number and see a full screen display of all interactors with this gene

   Doc → Click a number and see a full screen display of all articles about the gene

   Path → Click a number and see all pathways associated with the gene

   A GO annotation (Cellular Components, Biological Processes, Molecular functions) → Click a term and link out to the GO page for that term

   Navigation that is available in Query Interaction searches includes:

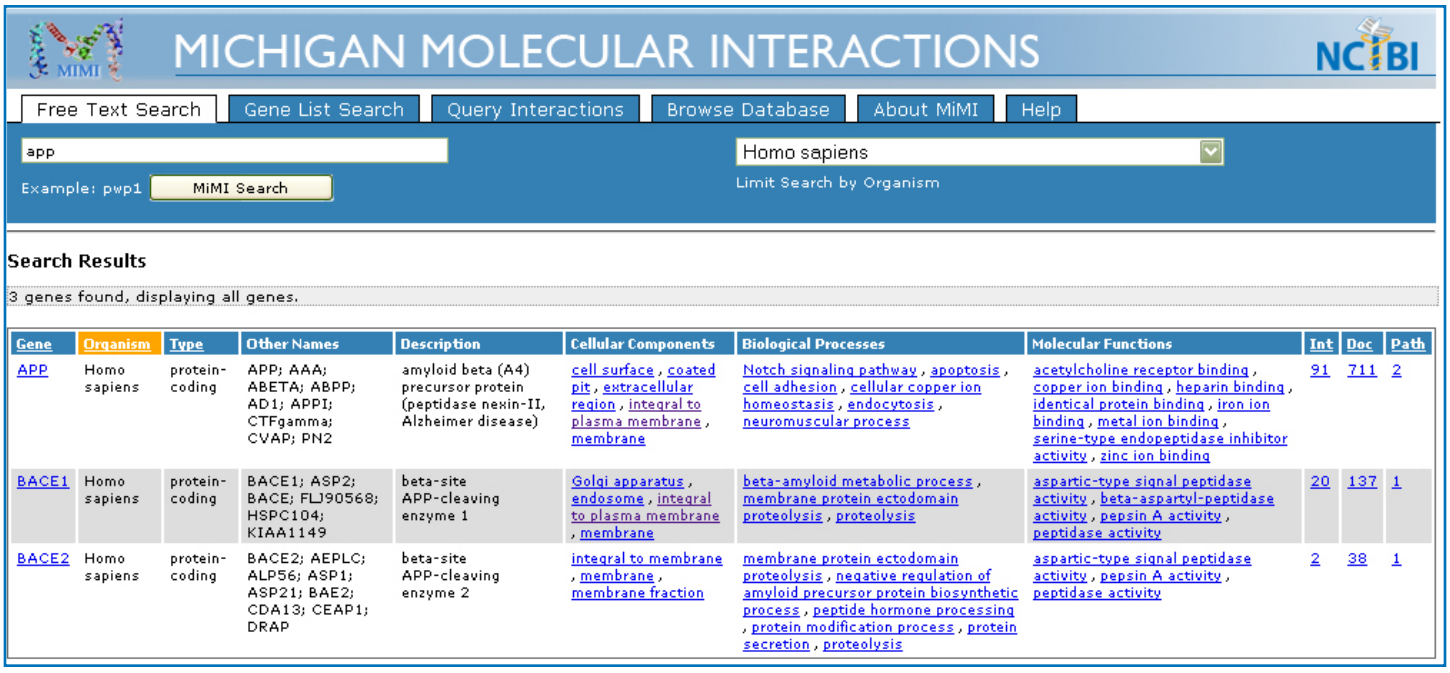
   Gene name 1 or Gene name 2 → Click the name to see the Gene Details page.

   Provenance → Click the gene name to see data from the mentioned source

   Lit Count → Click a number and see a full screen display of all related articles

2. Sort the Search Results table: Click on column headings that are underlined to sort in ascending or descending order. The sorted-on column will be highlighted after you click on it.

3. Save and export: Click on csv, Excel, or XML to save the entire Search Results table. If you navigate to interactors, articles, and pathways, you can also save these lists.



Gene	Organism	Type	Other Names	Description	Cellular Components	Biological Processes	Molecular Functions	Int	Doc	Path
APP	Homo sapiens	protein-coding	APP; AAA; ABETA; ABPP; AD1; APPI; CTFgamma; CVAP; PN2	amyloid beta (A4) precursor protein (peptidase nexin-II, Alzheimer disease)	cell surface, coated pit, extracellular region, integral to plasma membrane, membrane	Notch signaling pathway, apoptosis, cell adhesion, cellular copper ion homeostasis, endocytosis, neuromuscular process	acetylcholine receptor binding, copper ion binding, heparin binding, identical protein binding, iron ion binding, metal ion binding, serine-type endopeptidase inhibitor activity, zinc ion binding	91	711	2
BACE1	Homo sapiens	protein-coding	BACE1; ASP2; BACE; FLJ90568; HSPC104; KIAA1149	beta-site APP-cleaving enzyme 1	Golgi apparatus, endosome, integral to plasma membrane, membrane	beta-amyloid metabolic process, membrane protein ectodomain proteolysis, proteolysis	aspartic-type signal peptidase activity, beta-aspartyl-peptidase activity, pepsin A activity, peptidase activity	20	137	1
BACE2	Homo sapiens	protein-coding	BACE2; AEPLC; ALP56; ASP1; ASP21; BAE2; CDA13; CEAP1; DRAP	beta-site APP-cleaving enzyme 2	integral to membrane, membrane, membrane fraction	membrane protein ectodomain proteolysis, negative regulation of amyloid precursor protein biosynthetic process, peptide hormone processing, protein modification process, protein secretion, proteolysis	aspartic-type signal peptidase activity, pepsin A activity, peptidase activity	2	38	1

Figure 4.1 Search Results Overview Screen and Interactions You Can Perform.

4.2 Information found on the screen

Free Text and Gene List searches: The fields in the Search Results screens include:

Field Name	Description
Gene	The gene symbol or name.
Organism	The name of the organism the gene is found in.
Type	Gene Type. Several types of molecules exist within MiMI. These include: protein, DNA, RNA, polymer, complex, small molecule and photon.
Other Names	A list of synonyms (aliases) for this gene. Please note that this list is not complete, but merely the set of names found within all of our merged sources.
Description	Description content is created by curators of the source datasets. MiMI retains information from all sources, even if it is contradictory. In the case of descriptions, we will not repeat exact matches. However, since descriptions are normally English sentences, there are usually no exact matches and all versions are retained.
Cellular Components	GO annotations found in the original sources indicating where this molecule is found within the cell.
Biological Processes	GO annotations found in the original sources indicating all processes associated with this molecule.
Molecular Functions	GO annotations found in the original sources indicating this molecule's function within the cell.
Int	The number of interactors this gene has.
Doc	The number of related articles. Articles include those that have been curated and included in our merged databases as well as those retrieved through text mining.
Path	The number of pathways in which this gene appears.

Table 4.1 Description of the search results for free text and gene list searches.

Field Name	Description
Gene 1	The gene symbol or name.
Gene 2	The gene symbol or name or a blank field to see all interactors with Gene 1
Source Provenance	The sources from which data on the genes are derived.
Lit count	The number of articles referencing the given interaction.
Interaction info	Data drawn from the diverse sources integrated in MiMI characterizing the interaction, e.g. as protein-DNA, bi-directional,
Experiments	The source of evidence about an interaction. MiMI provides details on numerous types of experiments, allowing scientists to infer types of interactions and establish levels of confidence.

Table 4.2 Description of the search results for Query Interaction searches.

5.0 Analyzing and Interacting with Gene Details

5.1 Overview

You can navigate to comprehensive information about each gene of interest by clicking the Gene Name (hyperlinked) on the Search Results page.

Scanning and interacting with gene details facilitates the following tasks:

- Confirm that results in MiMI match results you expect to see
- Identify interactors and get detail on the interactions
- Look at literature and easily select, filter and save based on MeSH and other metadata
- Identify pathways containing the gene and other genes in each given pathway
- Link out to other biological resources, such as PFam
- Integrate with other NCIBI tools to extend explorations

The Gene Detail screen is arranged in 6 layers of information, described in detail in Information Found on the screen and summarized in Figure 5.1.

To show or hide a layer, click either the icon or "Show/Hide" in the heading.

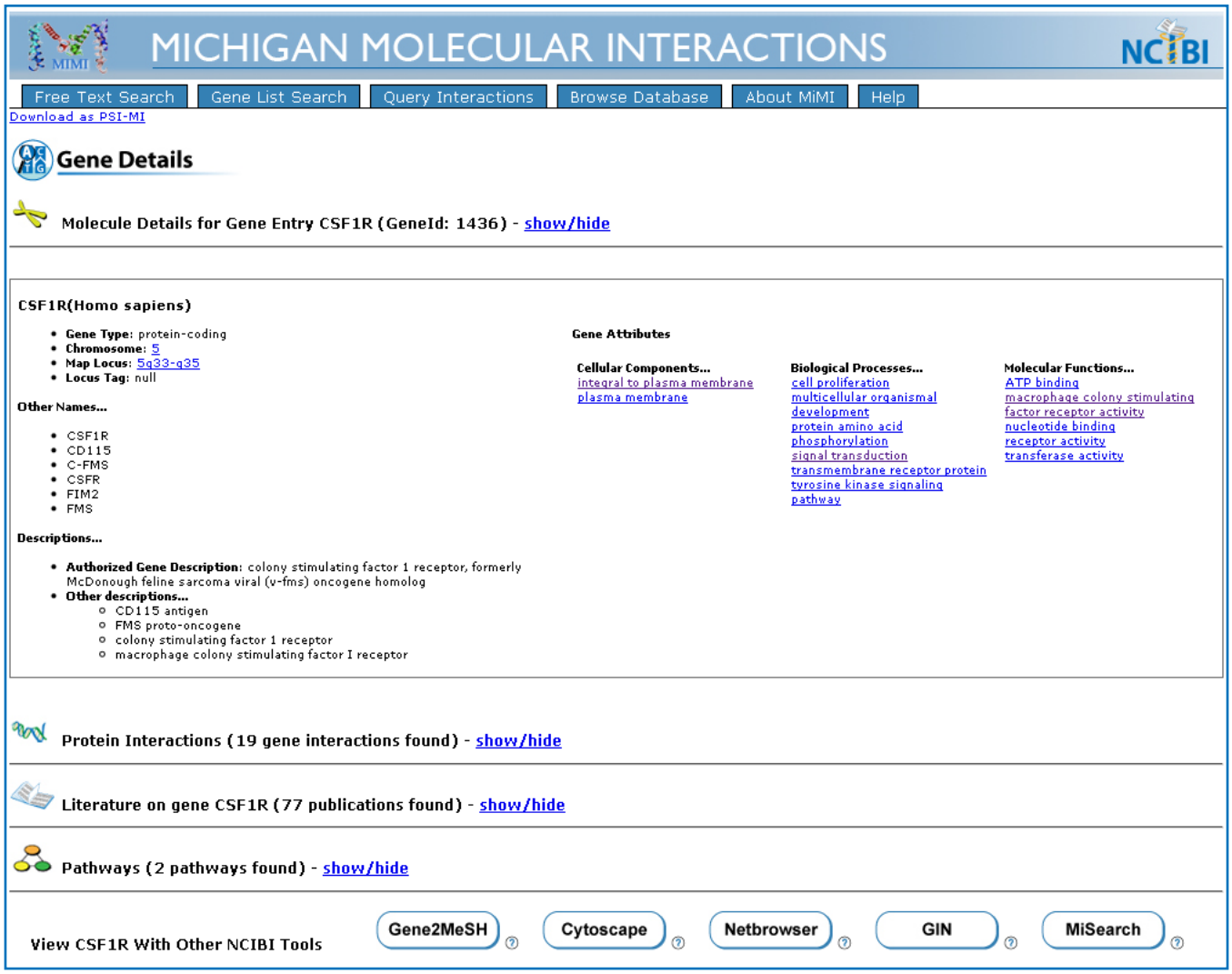


Figure 5.1 Details in each layer of the Gene Details Page.

5.2 Step by step procedures for navigating the page

1. Click either the heading icon or the Show/Hide tag to move to specific details within the screen and see the following:

Gene Information
See: Core information, e.g. name, aliases, ID, Description, Chromosome location

Link to: NCBI Map Viewer, NCBI Entrez Gene

Gene Attributes
See: GO annotations

Link to: GO site for a given annotation

Protein Interactions (number of protein-protein interactions)
See: Core information, e.g. interactor names, experiments producing the interaction, interaction description

Link to: Interactor details, PubMed citations about the interaction

Literature (number of PubMed articles that mention the gene)
See: Citation data, e.g. PMID, year, journal, title, author, abstract – drawn from merged databases (curated) and natural language processing (not curated)

Link to: Pubmed, an Interactive view of all articles on 1 page

Pathways (number of pathways in which the gene is a member)
See: Pathway names, descriptions – drawn from KEGG

Link to: KEGG details, KEGG pathway image, Other genes that belong to each pathway, Other genes that belong to all pathways combined

Links to Other Protein Information
See: Other available link-outs

Link to: PFam, OMIM, ReSeq, BLAST

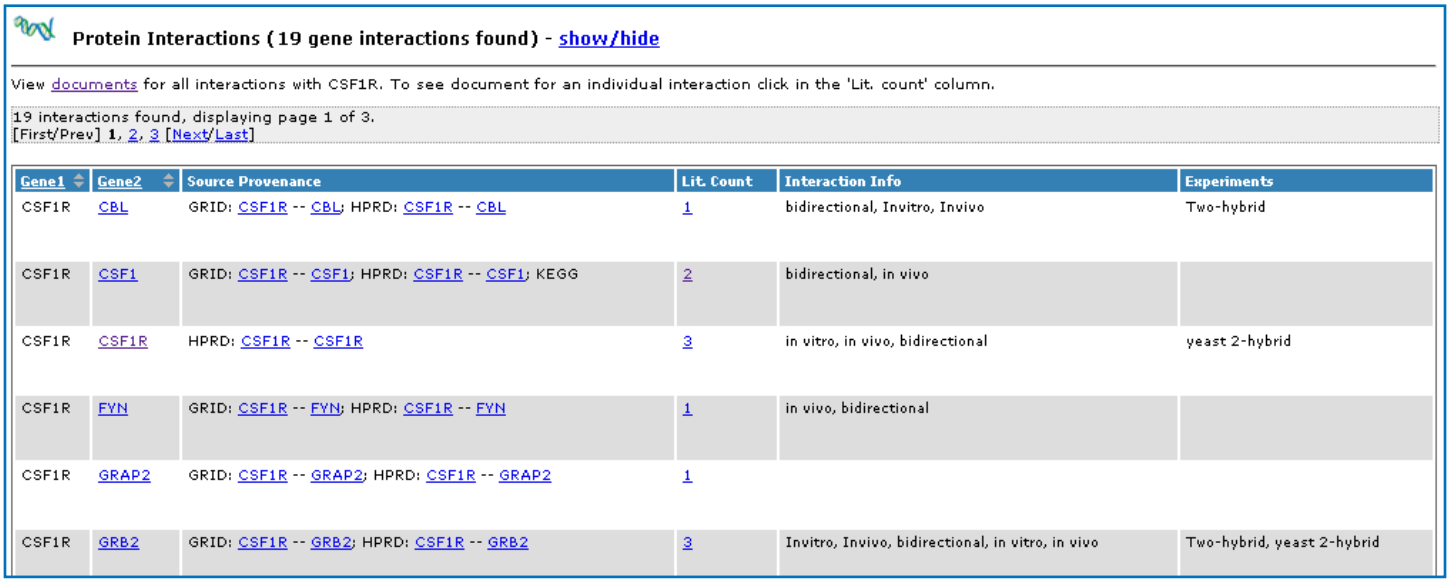
Integrated NCIBI Tools to Extend your Exploration
Clicking one of the NCIBI tool buttons will keep your place in your MiMI analysis while working with the other tool and the MiMI data:

Click	Your Purpose	Display Presents
Gene2Mesh	**Term enrichment and PubMed links.** Find overrep-resented MeSH for a gene and link out to related articles	MeSH terms that map to the focus gene, ranked by enrichment score. Also displays all genes enriched for a single MeSH term. Enables you to progressively explore a series of genes and MeSH terms and save/ download outcomes.
Cytoscape	**Network (visual) exploratory analysis and inference.** Interactively explore protein-protein interactions	A powerful interactive network showing the focus gene products, its neighbors, and their attributes. You can combine MiMI data with expression data, find shortest paths, recurrent subnetworks, and matches to KEGG pathways.
Netbrowser	**Network (visual) exploratory analysis.** Quickly view protein interactions and get a list of interactors.	A lightweight interactive network of protein interactions showing the focus gene product and its nearest neighbors. You can get details on proteins, download a list of proteins that interact with the focus gene, and expand one or more nodes (proteins) to see its neighbors.
GIN (Gene Interaction Network)	**Semantic summaries and PubMed links.** View text summaries mined from PubMed about the focus gene and its interactors	Summaries extracted from PubMed articles through semantic natural language processing methods. Also displays citation information, and links to networks of interactions.
MiSearch	**PubMed queries through an adaptive interface** Search PubMed and quickly refine queries without writing compound SQL statements.	Search results for the focus gene from a PubMed query with a means to quickly refine results through easy-to-formulate compound queries with MeSH terms. Stores selected articles of interest in a Profile and uses them to determine the relevance of results in subsequent searches and order accordingly.

2. Click on a link in any of the detail sections to move to a different workspace.

5.3 Step by step procedures for examining interactions

1. Click the icon or Show/Hide tag in the Interaction section. A table of interacting proteins associated with genes appears.



Protein Interactions (19 gene interactions found) - show/hide

View documents for all interactions with CSF1R. To see document for an individual interaction click in the 'Lit. count' column.

19 interactions found, displaying page 1 of 3.
[First/Prev] 1, 2, 3 [Next/Last]

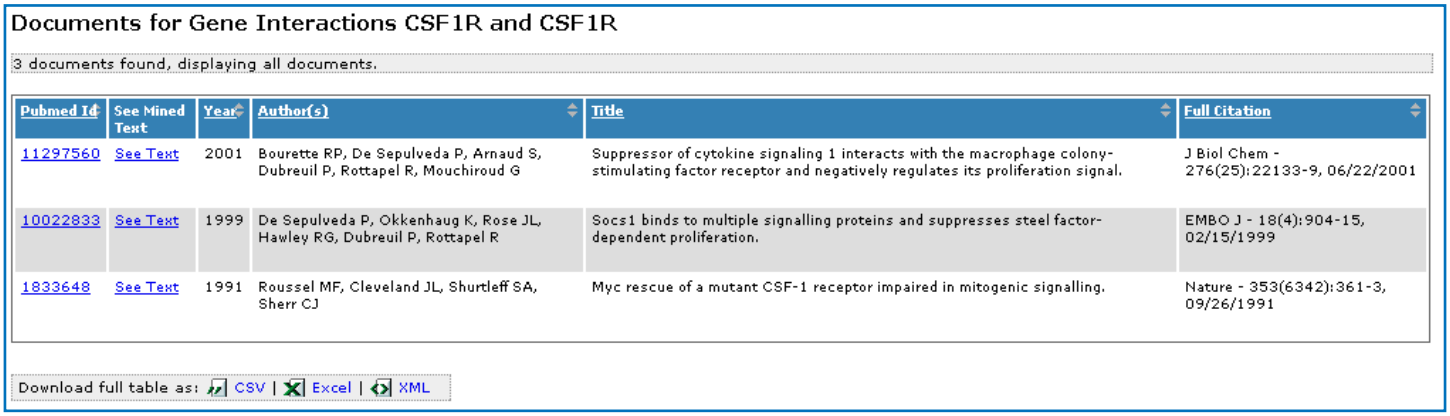
Gene1	Gene2	Source Provenance	Lit. Count	Interaction Info	Experiments
CSF1R	CBL	GRID: CSF1R -- CBL; HPRD: CSF1R -- CBL	1	bidirectional, Invitro, Invivo	Two-hybrid
CSF1R	CSF1	GRID: CSF1R -- CSF1; HPRD: CSF1R -- CSF1; KEGG	2	bidirectional, in vivo	
CSF1R	CSF1R	HPRD: CSF1R -- CSF1R	3	in vitro, in vivo, bidirectional	yeast 2-hybrid
CSF1R	FYN	GRID: CSF1R -- FYN; HPRD: CSF1R -- FYN	1	in vivo, bidirectional	
CSF1R	GRAP2	GRID: CSF1R -- GRAP2; HPRD: CSF1R -- GRAP2	1		
CSF1R	GRB2	GRID: CSF1R -- GRB2; HPRD: CSF1R -- GRB2	3	Invitro, Invivo, bidirectional, in vitro, in vivo	Two-hybrid, yeast 2-hybrid

Figure 5.2 Display of Interactions.

2. Navigate this section as follows:

   - Sort by column: Click Gene 2 in the column heading to sort alphabetically by interactor. The whole list (including what is not immediately visible) gets reordered.
   - Click the name of any genes in the Gene 2 column to go to its Gene detail page
   - Click csv, Excel or XML to download the citations for a specified interaction.
   - Click the number in the Literature Count column that indicates the number of articles about the given interaction. The full citation(s) appears.

3. Interact with the literature about interactions after clicking the number in Lit Count.



Documents for Gene Interactions CSF1R and CSF1R

3 documents found, displaying all documents.

Pubmed Id	See Mined Text	Year	Author(s)	Title	Full Citation
11297560	See Text	2001	Bourette RP, De Sepulveda P, Arnaud S, Dubreuil P, Rottapel R, Mouchiroud G	Suppressor of cytokine signaling 1 interacts with the macrophage colony-stimulating factor receptor and negatively regulates its proliferation signal.	J Biol Chem - 276(25):22133-9, 06/22/2001
10022833	See Text	1999	De Sepulveda P, Okkenhaug K, Rose JL, Hawley RG, Dubreuil P, Rottapel R	Socs1 binds to multiple signalling proteins and suppresses steel factor-dependent proliferation.	EMBO J - 18(4):904-15, 02/15/1999
1833648	See Text	1991	Roussel MF, Cleveland JL, Shurtleff SA, Sherr CJ	Myc rescue of a mutant CSF-1 receptor impaired in mitogenic signalling.	Nature - 353(6342):361-3, 09/26/1991

Download full table as: CSV | Excel | XML

Figure 5.3 Literature about Interactions.

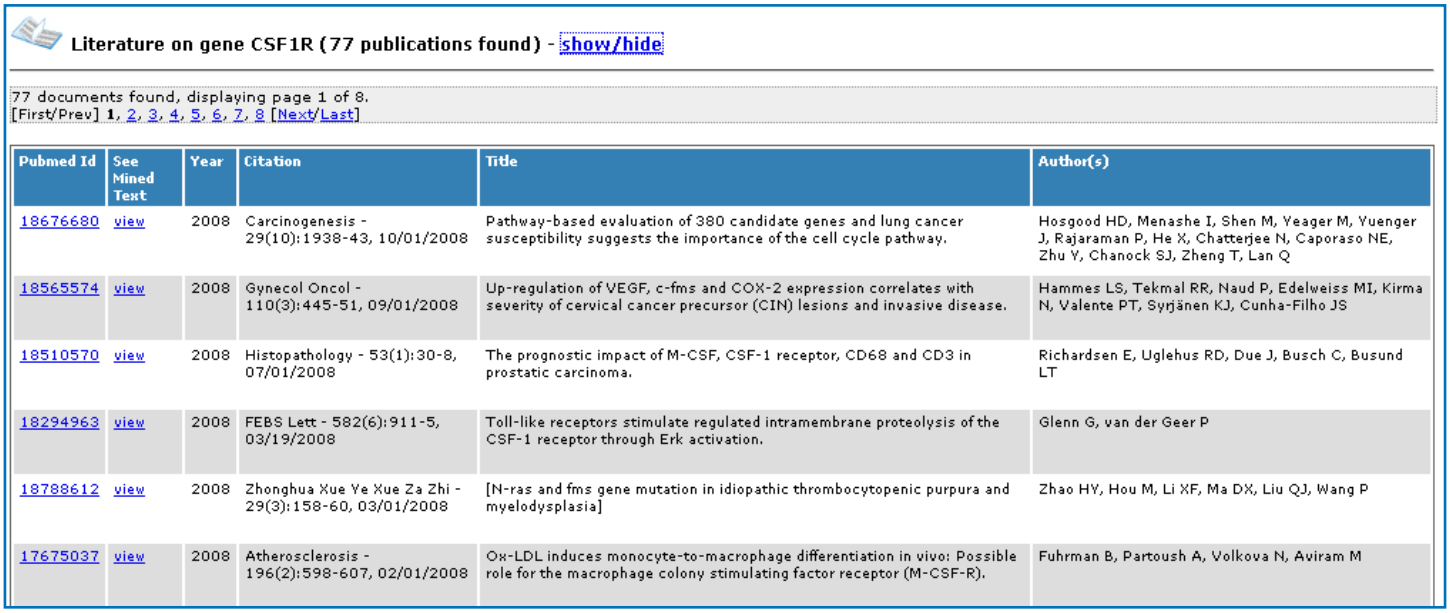
In the Literature page:

- Click any column heading to sort the table by the column.
- Click the PMID in any row to see the full PubMed citation of the given article.
- Under the Mined Text column click a row's link to read extracts mentioning interactions. A screen of summarized text appears.
- Click csv, Excel, or XML to save and/or download the displayed literature.
- Press your browser's back button to return to the MiMI Interactions page.

5.4 Step by step procedures for examining literature

Literature displayed in MiMI is drawn from curated sources, from PubMed searches, and from natural language processing (NLP). See 2.2. Curation and Sources for more details.

1. Click the icon or the Show/Hide tag in the Literature heading. The section opens and displays articles that reference the focus gene (See Figure 5.4). The display shows the first page of many pages of articles if numerous articles refer to the gene/gene products.



Literature on gene CSF1R (77 publications found) - show/hide

77 documents found, displaying page 1 of 8.
[First/Prev] 1, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

Pubmed Id	See Mined Text	Year	Citation	Title	Author(s)
18676680	view	2008	Carcinogenesis - 29(10):1938-43, 10/01/2008	Pathway-based evaluation of 380 candidate genes and lung cancer susceptibility suggests the importance of the cell cycle pathway.	Hosgood HD, Menashe I, Shen M, Yeager M, Yuenger J, Rajaraman P, He X, Chatterjee N, Caporaso NE, Zhu Y, Chanock SJ, Zheng T, Lan Q
18565574	view	2008	Gynecol Oncol - 110(3):445-51, 09/01/2008	Up-regulation of VEGF, c-fms and COX-2 expression correlates with severity of cervical cancer precursor (CIN) lesions and invasive disease.	Hammes LS, Tekmal RR, Naud P, Edelweiss MI, Kirma N, Valente PT, Syrjänen KJ, Cunha-Filho JS
18510570	view	2008	Histopathology - 53(1):30-8, 07/01/2008	The prognostic impact of M-CSF, CSF-1 receptor, CD68 and CD3 in prostatic carcinoma.	Richardsen E, Uglehus RD, Due J, Busch C, Busund LT
18294963	view	2008	FEBS Lett - 582(6):911-5, 03/19/2008	Toll-like receptors stimulate regulated intramembrane proteolysis of the CSF-1 receptor through Erk activation.	Glenn G, van der Geer P
18788612	view	2008	Zhonghua Xue Ye Xue Za Zhi - 29(3):158-60, 03/01/2008	[N-ras and fms gene mutation in idiopathic thrombocytopenic purpura and myelodysplasia]	Zhao HY, Hou M, Li XF, Ma DX, Liu QJ, Wang P
17675037	view	2008	Atherosclerosis - 196(2):598-607, 02/01/2008	Ox-LDL induces monocyte-to-macrophage differentiation in vivo: Possible role for the macrophage colony stimulating factor receptor (M-CSF-R).	Fuhrman B, Partoush A, Volkova N, Aviram M

Figure 5.4 Literature about a focus gene.

2. To navigate this screen:

   - Click the number or "Next" link at the top of the table to move through pages of literature.
   - Click the PubMed ID (PMID) to go directly to the full PubMed citation. You can download from there, move to the full text article, and see the PubMed listing of related articles.
   - Click csv, Excel, or XML to download the portion of the literature list that is displayed on the screen.
   - Click "View" under See Mined Text to read passages mined from the article about the gene and other genes associated with it. The Annotated Article Abstract screen appears (See Figure 5.5)

3. On the Annotated Article Abstract screen:

   - Move to the PubMed screen for the article by clicking the PMID.
   - Quickly find the focus gene and genes associated with it through the highlighting in the abstract.
   - See if highlighted association comes from tagging or merging by referring to the Gene Detail table
   - Move to more detail on an associated gene by interacting with the Gene Detail table.
   - Click csv, Excel, or XML to download the Gene Detail table on this page.
   - Press the Back Arrow to return to your originating MiMI Gene Detail page.

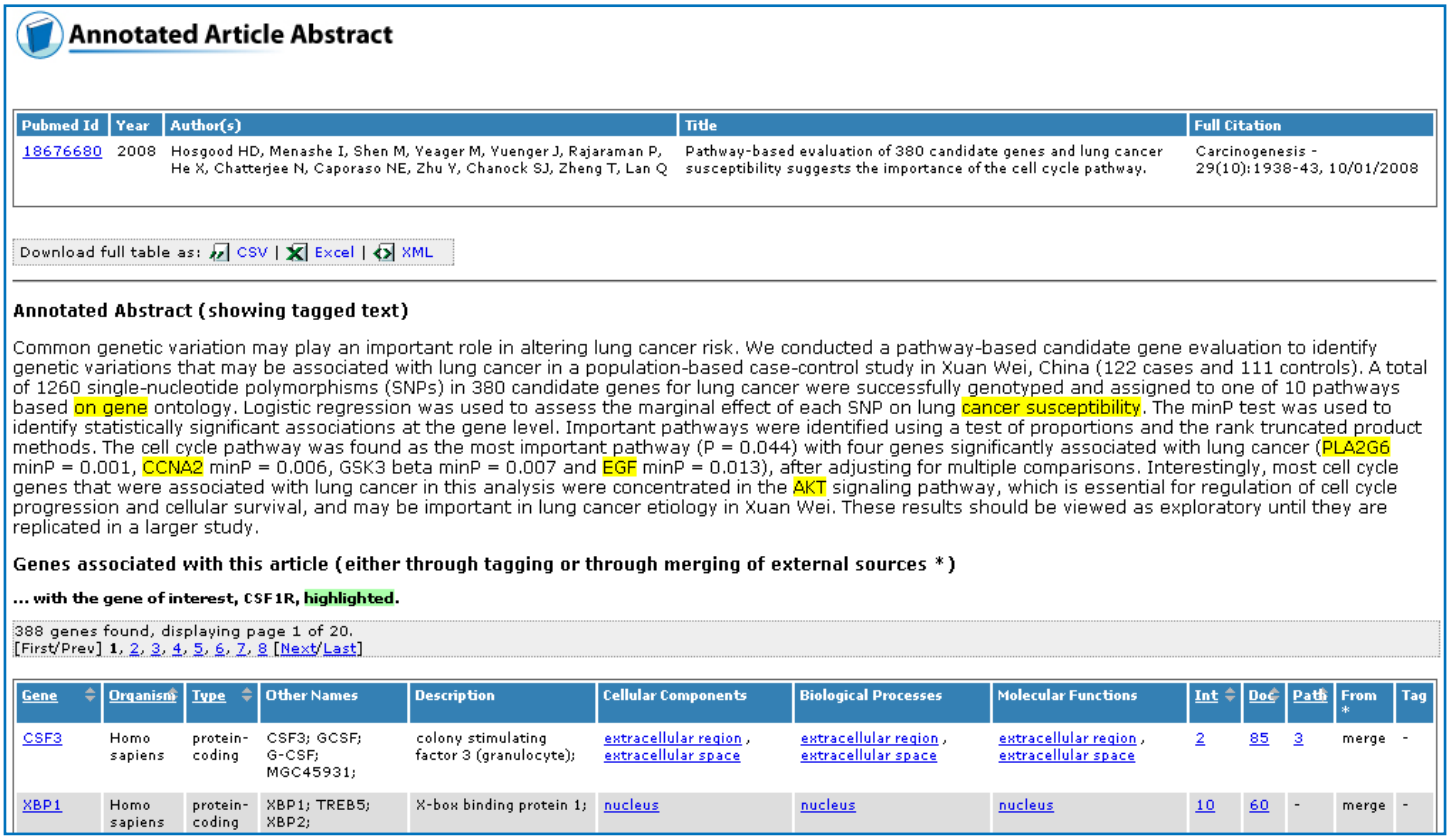


Figure 5.5 Annotated Article Abstract screen.

5.5 Step by step procedures for examining pathways

Pathways related to genes are drawn from the KEGG and Reactome databases.

1. Click the icon or the Show/Hide tag in the Pathways heading. The section opens and displays pathways related to the gene (See Figure 5.6).

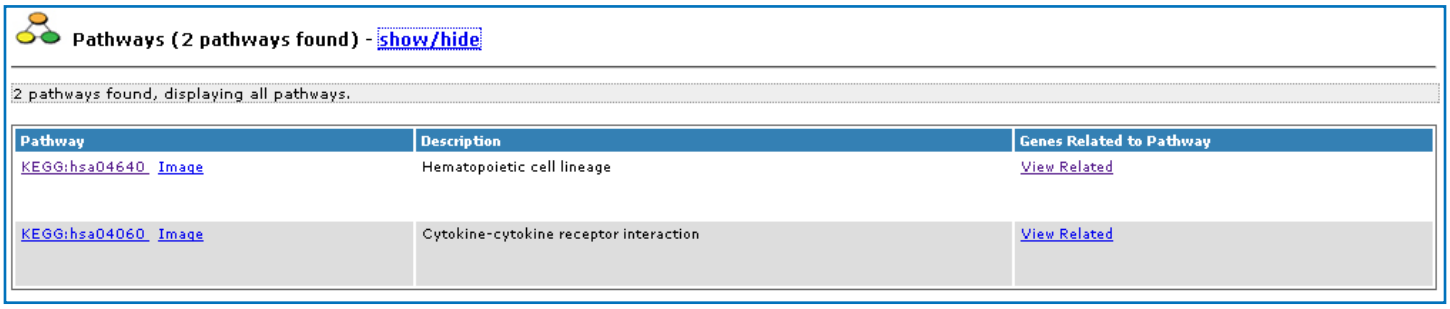


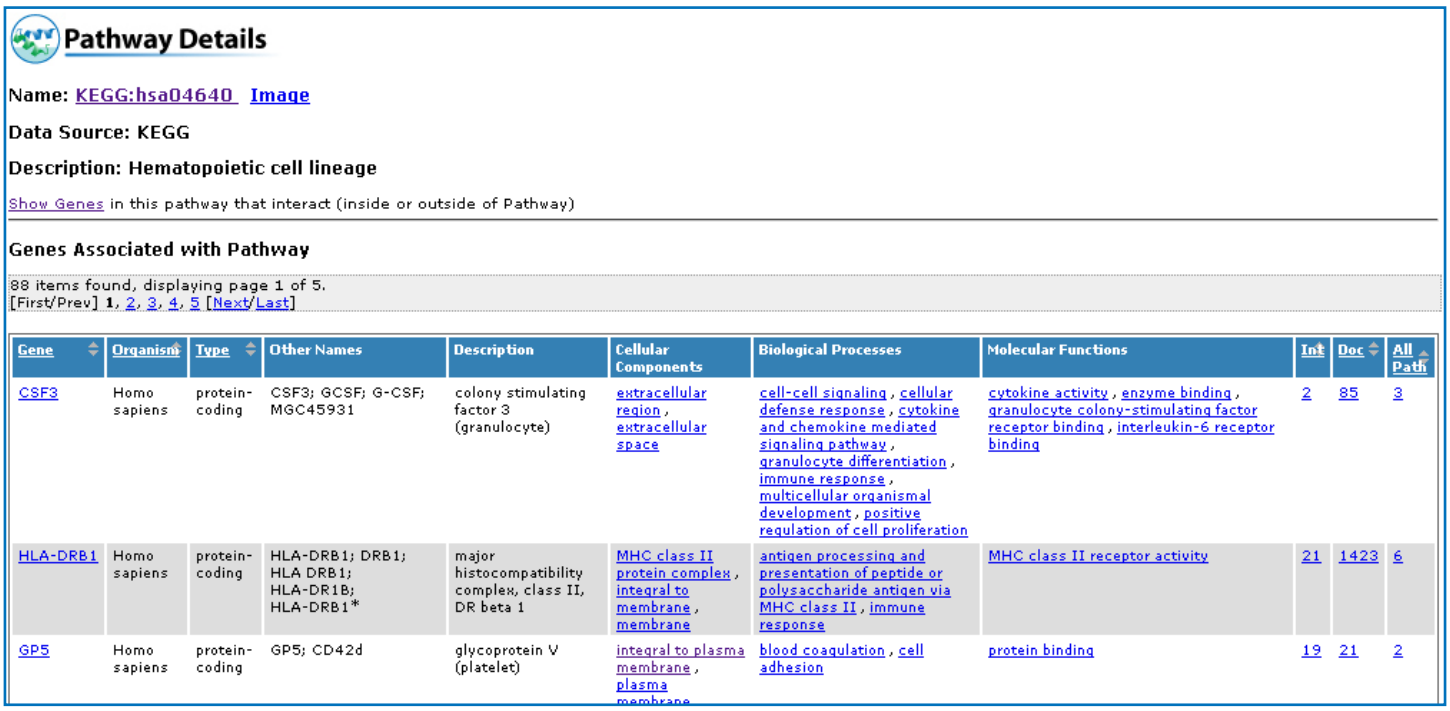
Figure 5.6 Pathways related to a focus gene.

2. To navigate this screen:

   - Click the "KEGG:…" link to go to the KEGG site description of this pathway.
   - Click "Image" to see a network view of the pathway
   - In the column Genes Related to Pathway, click "View Related" to see a list of all genes in the pathway. This list may take time to load.

3. On the Genes Related to Pathway screen (Figure 5.7):

   - Click the KEGG links to see a pathway description or network image.
   - Move through the table of genes by clicking the page numbers or Prev/Next.
   - Follow 4.1. Search Results screen procedures above to navigate the gene table.
   - Click csv, Excel, or XML to download the Gene Detail table on this page.
   - Press the back button in your browser to return to your originating MiMI Gene Detail page.
   - Click "Show Genes in this pathway that interact (inside or outside of Pathway)" to see a list of all interacting gene. This list may take time to load.



Pathway Details

Name: KEGG:hsa04640 Image

Data Source: KEGG

Description: Hematopoietic cell lineage

Show Genes in this pathway that interact (inside or outside of Pathway)

Genes Associated with Pathway

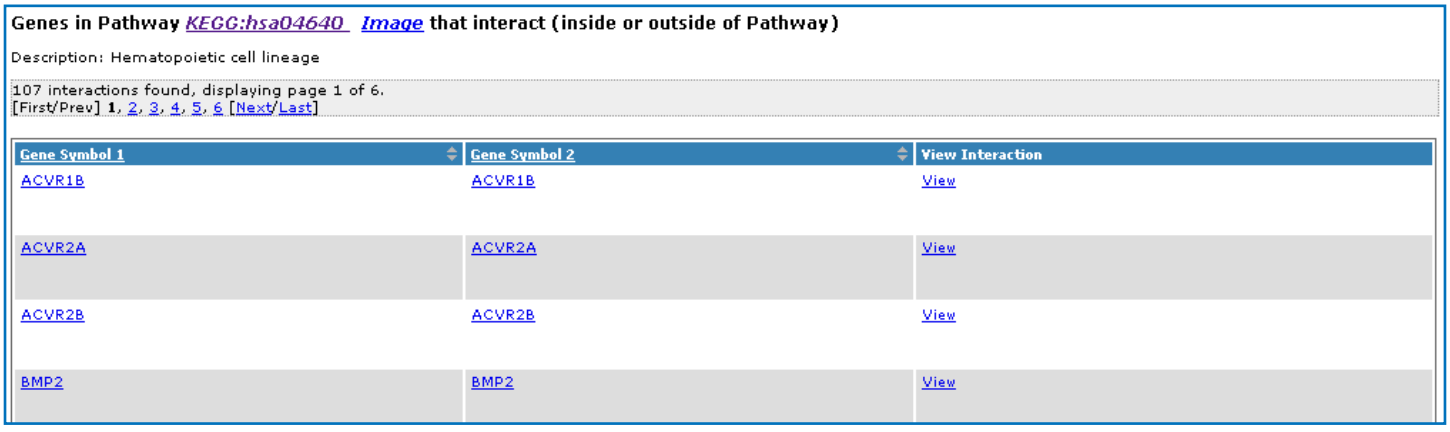
88 items found, displaying page 1 of 5.
[First/Prev] 1, 2, 3, 4, 5 [Next/Last]

Gene	Organism	Type	Other Names	Description	Cellular Components	Biological Processes	Molecular Functions	Int	Doc	All Path
CSF3	Homo sapiens	protein-coding	CSF3; GCSF; G-CSF; MGC45931	colony stimulating factor 3 (granulocyte)	extracellular region, extracellular space	cell-cell signaling, cellular defense response, cytokine and chemokine mediated signaling pathway, granulocyte differentiation, immune response, multicellular organismal development, positive regulation of cell proliferation	cytokine activity, enzyme binding, granulocyte colony-stimulating factor receptor binding, interleukin-6 receptor binding	2	85	3
HLA-DRB1	Homo sapiens	protein-coding	HLA-DRB1; DRB1; HLA DRB1; HLA-DR1B; HLA-DRB1*	major histocompatibility complex, class II, DR beta 1	MHC class II protein complex, integral to membrane, membrane	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II, immune response	MHC class II receptor activity	21	1423	6
GP5	Homo sapiens	protein-coding	GP5; CD42d	glycoprotein V (platelet)	integral to plasma membrane, plasma membrane	blood coagulation, cell adhesion	protein binding	19	21	2

Figure 5.7 Genes Related to Pathway screen.

4. On the Genes in Pathway that Interact screen (Figure 5.8):

   - See how many interactions there are and move through the pages as needed.
   - Sort by Gene 1 or Gene 2 by clicking the column heading.
   - Click the gene names to move to a gene detail table on the selected gene.
   - Click "View Interaction" to move to an interaction detail table on the interaction.
   - Press the back button in your browser to return to your originating MiMI Gene Detail page



Genes in Pathway KEGG:hsa04640 Image that interact (inside or outside of Pathway)

Description: Hematopoietic cell lineage

107 interactions found, displaying page 1 of 6.
[First/Prev] 1, 2, 3, 4, 5, 6 [Next/Last]

Gene Symbol 1	Gene Symbol 2	View Interaction
ACVR1B	ACVR1B	View
ACVR2A	ACVR2A	View
ACVR2B	ACVR2B	View
BMP2	BMP2	View

Figure 5.8 Genes in Pathway that Interact screen.

5.6 Step by step procedures for linking out to other sources

MiMi provides quick links to external sites that have proven to be valuable to biomedical researchers.

1. Click on the links provided: Pfam, OMIM, RefSeq, or BLAST. MiMI takes you to the relevant pages in the external resource.
2. Return to MiMI by pressing the back button in your browser.

5.7 Step by step procedures for linking to integrated NCIBI tools

Additional tools are integrated into the MiMI application for you extend your analysis of gene interactions. Clicking on the ? button next to each individual NCIBI tool button will give you additional information about that tool.

5.7.1 Gene2MeSH

If you wish to find the most statistically significant Medical Subject Headings for a gene of interest and see related articles from PubMed then:

1. Click the Gene2MeSH button. A new Gene2MeSH web browser window will open with the dynamically generated search results of your specific gene name limited to Humans for the organism type.

   MeSH terms will be listed in order of statistical significance (Fisher's Exact column).

   The MeSH Category column lists the specific subheading classification assigned to that MeSH term when an article is indexed in PubMed. Not all MeSH terms will have a subheading classification designated; in which case that particular MeSH Category cell will be empty.

   The Gene Description column lists the full name from the National Center for Biotechnology Information's Entrez Gene database.
2. Click the hyperlinked PubMed Articles number to view the list of articles that contain the gene name/MeSH pairing. The first article will be displayed in PubMed in a new web browser window.
3. Click the About link for additional information (features, searching, downloading results) about Gene2MeSH.

5.7.2 Cytoscape

Cytoscape is an open source software platform for visualizing molecular interaction networks and allows you to integrate expression and other state data.

1. Click the Cytoscape button. Cytoscape launches with a Run Java screen. It displays a dialogue box that asks you to run the application. Press Run.
2. The Cytoscape-generated network for your focus gene and its nearest neighbors is displayed. Your focus gene is the triangular node; neighbors are circles. The network is in an unselected state.
3. Sweep the network with your mouse, press Ctrl+ Alt+A, or Go to the Select menu → Select All Nodes and Edges, to highlight the select the network. It becomes highlighted.
4. Identify the items in the network by going to the data panel (lower panel) and clicking the far left data table icon. A pop up checklist appears with field names that you can display.
5. Check the field names in which you are interested. Common fields for initial views are Component, Gene Name, Function, Other Gene Names, Pathway, and Process. Data relevant to the network for these fields now fill in the data panel table.
6. Move columns around in the table and sort by clicking the column headings to get the table into an order that is easy to read and analyze.
7. Use the magnifying glasses to zoom in and out of the network. The lower image in the left panel lets you navigate zoomed views. Move the box to navigate.
8. Click any row in the data panel table and the corresponding node lights up in green.
9. Click any node or sweep many nodes directly on the network and the selection highlights in yellow. The data panel table adjusts to show data on only the selected items.
10. Refer to the Cytoscape Help and user materials and to the documentation for the MiMI plug to Cytoscape for more instruction on exploring protein interactions visually.

5.7.3 Netbrowser

Netbrowser is a molecular interaction network visualization tool that displays your gene product and its nearest neighbors.

1. Click the Netbrowser button. The web page will refresh to Netbrowser displaying a network of genes that directly interact with your focus gene. Nodes are labeled with the gene names.
2. Customize the network display by clicking on an individual node and moving it on the screen.

   Click on the slider(s) for Motion-Stop Threshold and Repulsion Factor to adjust how the nodes interact with each other on the screen as individual nodes are moved around.

   Click on the Reset Graph button to reset the network display to its original form.
3. Double click on an individual node to display the gene details of that gene.
4. Click on the Gene List button to view a tabular list of all the genes displayed in the network on the screen.

5.7.4 GIN

GIN, Gene Interaction Network, is a tool that uses natural language processing methods to mine text for information to allow you to browse articles and molecular interaction information.

1. Click the GIN button. The web page will refresh to display interaction results from the Molecule Search in GIN. If GIN does not have any molecule matches, then the page will display "No molecule matches the given query".

   Interaction results will display the first neighbor gene names and an excerpt from the relevant article with gene names highlighted.

   If second neighbor information is available, it will also be displayed.

2. Click on the gene name to run a molecule search on that gene name.
3. Click on the article ID to display the full text of the article with all molecule interaction relevant sentences highlighted. A link to the PubMed Central version of the article is also provided.
4. Click the Home link for additional information on GIN.

5.7.5 MiSearch

MiSearch is an adaptive PubMed search tool that creates a history of your browsing profile to create a ranked list of citations from PubMed that displays articles that it believes will be most relevant to you at the top of the list.

1. Click the MiSearch button. The web page will refresh to display search results of your gene name in the Query field of MiSearch. PubMed articles matching your criteria will be listed in a ranked order on the page.
2. Mark articles of interest to you by checking the boxes next to the PubMed ID and click the Submit button again to have the articles reordered and reranked so that those that were marked by you appear near the top of the list.
3. Click on the Show/Hide links pertaining to the Key Authors, MeSH Terms and Compounds sections to extend your search. After the expanded lists are displayed, you can click on the + signs next to individual authors, terms or compounds to add those items to your search strategy.
4. Click on the MiSearch Help link for additional information (background, searching tips) about MiSearch.

5.8 Information found on the screen

Some fields on this screen may not be populated. They are blank if none of our curated sources has any information for a field such as cellular component

Gene Details	Description
Chromosome	The chromosome on which this gene is related to. (When known)
Descriptions	Authorized Gene Description Other Gene Description Molecule Description
Gene Type	Type of Gene
Literature on Gene	Includes PubMed ID (PMID) for all articles related to the given gene along with: Year of publication, Journal name, title, and author. Includes a link to MiMI Interactive Literature, which lets you see all articles in a single list, sort them, and selectively save/export those of interest.
Map Locus	The chromosome location of the gene.
Organism	The name of the organism the gene is found in. If we do not recognize a taxonomy id, this field will contain "Unspecified".
Other Names	This field will contain the list of synonyms for this gene. Please note that this list is not complete, but merely the set of names found within all of our ingested sources.
Gene Attributes	**Description**
Biological Processes	Any GO terms found in the original sources indicating all processes this gene's products are associated with.
Cellular Components	Any GO terms found in the original sources indicating where this gene's products are found within the cell.
Molecular Functions	Any GO terms found in the original sources indicating this gene's products' function within the cell.

Protein Interactions	Description
Experiment	The source of evidence about an interaction. MiMI provides details on numerous types of experiments, allowing scientists to infer types of interactions and establish levels of confidence.
Gene Prod. 1	The name of the protein related to the focus gene (the gene on which you have selected detailed information. Clicking the gene name link will take you to the Gene Detail Page for the gene.
Gene Prod. 2	The name of the protein with which there is an interaction. Clicking the gene name link will take you to the Gene Detail Page for the gene.
Interaction Information	Many interactions in MiMI are not 'complete'. They may lack conditions, interaction sites, etc. Interactions in MiMI are the sum of information from the data and attribute sources. If no condition information is contained in any of these sources for this interaction, then MiMI will not contain any condition information. Likewise, if no interaction site information is available from the source datasets, the MiMI molecule will not contain an interaction site field. Several types of interactions exist within MiMI. These include: bidirectional, acts on, acted_upon, and complex. With regard to First and Second Interactor, it should be read as "Gene 1 acts on Gene 2". Experimental, classification, and NLP derived interaction information.
Lit Count	The number of documents that reference this interaction
Source Provenance	Lists provenance sources for the interaction – see Preface. The original data source from which the information came.
Literature	**Description**
Author(s)	The authors of the publication.
Citation	The citation.
Pubmed Id	Lists all of the PubMed IDs (PMIDs) that were cited in the original data sources for these genes. Clicking on the PMID link will take you directly to the citation in the PubMed database.
Search Google Scholar	Links to results from a Google Scholar search on the gene
Title	The title of the publication.
View → Annotated Article Abstract	Links to part of the abstract that references the gene in relation to other genes, when available, and gives details on the other genes
Year	The year of the publication.

Pathways	Description
Description	The name of the pathway.
Pathway	The KEGG Pathway database entry. The KEGG Pathway database is a collection of graphical diagrams (KEGG pathway maps) and associated text information (KEGG pathway entries). Each pathway is identified by a five-digit number, preceded by ko, map, rn, and three- or four-letter organism code, distinguishing reference pathways and organism-specific pathways. Clicking the hyperlink takes you to the KEGG Pathway database specific entry.
Pathways Details → Genes Associated with Pathway	A link to the Pathway Details page which describes genes and molecules associated with the pathway. Clicking on the View link will take you to the Pathway Details page.
Show Genes → Genes that Interact in the Pathway	A link to details about genes found in the pathway that (a) interact with each other and (b) products of genes found in the pathway with known interactors outside the pathway, as stored in MiMI.
Link outs	**Description**
Pfam	Pfam database: a collection of protein families represented by multiple sequence alignment and hidden Markov models.
OMIM	A database of known diseases with a genetic component with information on relevant genes, when known
RefSeq	NCBI database on nucleotide sequences and their protein translation
BLAST	Basic Local alignment Search Tool for comparing a query biological sequence with a library or database of sequences to find similarities
Export/Save	**Description**
csv	Comma separated value format
Excel	Microsoft Excel format
PSI-MI	Protein to protein data exchange format
XML	Extensible Markup Language format