National Center for Integrative Biomedical Informatics

Suite of Tools

University of Michigan

NCIBI Resources

NATIONAL CENTER FOR INTEGRATIVE BIOMEDICAL INFORMATICS

Michigan Molecular Intreactions Index (MiMI)

Protein interaction data mining tool

Access to this Tool: **http://mimi.ncibi.org**

Protein interaction data exists in a number of repositories. Much like the tale of the blind men and the elephant, each repository presents a limited view of the available molecular interaction information to the biomedical researcher. Each source is distinct in its biological focus, data format and user interface(s). To review all the available data, a researcher must visit each repository individually – an information gathering process that can be both tedious and time consuming.

MiMI strives to benefit the research community by gathering and integrating data from well-known protein interaction databases using an intelligent deep-merging strategy to present a unified view based on the concept of a "MiMI molecule". The "MiMI molecule" attempts to represent the concept of a molecular entity as it is used in the biomedical research community. With heuristic, well-validated identity functions, molecules that may have different identifiers but represent the same entity are merged. To help the researcher judge the validity of the aggregated data, the provenance of each piece of data is tracked during the data integration process. Complementary and contradictory data can be traced back to the original source for review and evaluation.

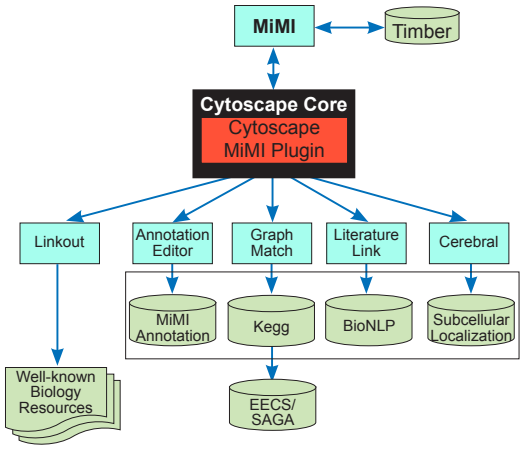
Today MiMI is being utilized by tools within NCIBI such as PubVIZ, extended to work with other tools such as Cytoscape and made directly available to researchers. MiMI serves the needs of a diverse community by making its data accessible in open and standard formats including XML, Web and SOAP access.

Contact for MiMI:
gtarcea@umich.edu

A Cytoscape Plugin to Access the MiMI Molecular Interactions Index

Access to this Tool: **http://www.bioinformatics.med.umich.edu/app/nlp/logo/pluginHelp/pluginHelp.html**

Cytoscape is a widely used open source software tool for the analysis of biomolecular interaction networks. The power of Cytoscape is most apparent when it is coupled to databases of protein-protein, protein-DNA, and genetic interactions. Protein interaction data exists in many repositories, each with its own data format, molecule identifier and supplementary information. Michigan Molecular Interactions Index (MiMI) integrates data from multiple well-known protein interaction databases using an intelligent deep-merging approach. The two resources complement each other well. To facilitate access to the molecular interaction data assembled in MiMI from the Cytoscape user interface, and to allow biological users to readily explore and analyze interaction data from MiMI, we have implemented a Cytoscape plugin that communicates with the MiMI database.



Contact for Cytoscape Plugin:
jinggao@umich.edu

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
3600 Green Court, Suite 700, Ann Arbor, Michigan 48105-1570

Phone: 734-615-5914
Fax: 734-998-8571
http://www.ncibi.org/

NCIBI Resources

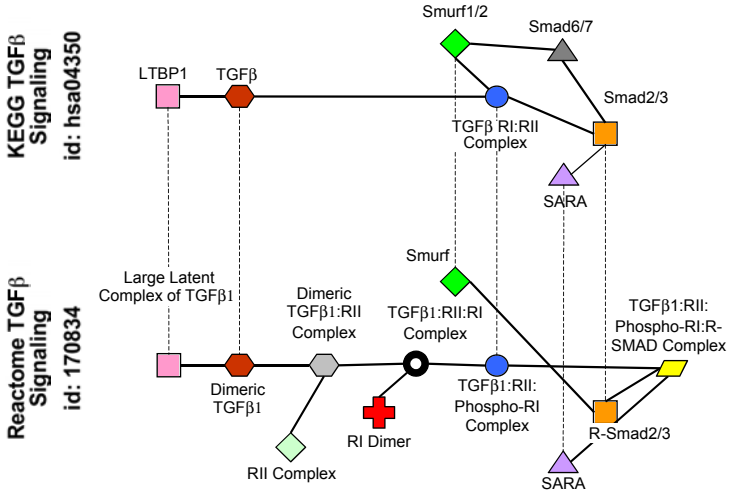
NATIONAL CENTER FOR INTEGRATIVE BIOMEDICAL INFORMATICS

Substructure Index Based Approximate Graph Alignment (SAGA)

A fast and flexible graph matching tool

Access to this Tool: **http://enigma.eecs.umich.edu/saga.html**

SAGA is a tool for querying biological graph database. SAGA implements an efficient approximate subgraph matching algorithm that can be use for a variety of biological graph matching problems. Examples, of such applications are: pathway matching, and matching literature databases that have been parsed into semantic graphs. In the above example SAGA is used to compare pathways in KEGG and Reactome. SAGA is able to find matches for similar information in these two different databases even though they organize pathways in different ways. Note comparing pathways from different databases can be a precursor to pathway data integration. In this second example, SAGA is used to query a literature graph database. For this example, portions of PubMed were parsed into graphs that have nodes representing gene names. A link is drawn between two genes if they are discussed in the same sentence (indicating there is potential association between the two genes).



Contact for SAGA:

Jignesh Patel jignesh@eecs.umich.edu

Yuanyuan Tian ytian@eecs.umich.edu

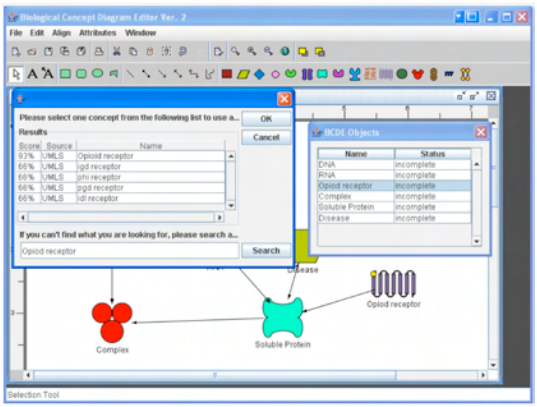
Biological Concept Diagram Editor (BCDE)

Molecular interactions, bioinformatics tools, workflows, Diagram Editor

Access to this Tool: **http://arrayanalysis.mbni.med.umich.edu/draw/**

Biological Concept Diagram Editor (BCDE) is a conceptual relationship diagramming tool specifically designed for biomedical researchers. It allows for efficient knowledge/data capture, fast diagram creation, easy data retrieval and flexible exporting.

The BCDE application is based on the JHotDraw framework. Through it, users can create, modify, load, and save BCDE diagrams. Each BCDE figure can be annotated using fields from the BioPAX level II format. In addition, a user can add URL links and attachments to a BCDE figure. A major design goal of BCDE is to greatly increase efficiency of capturing data from electronic sources; to this end we develop a transfer function to allow for drag-and-drop input methods with text, images, and files. In addition to transferring user selected input, we also include certain metadata such as the location address (URL) of the page currently displayed and time/date of the transfer. Diagrams generated in BCDE are stored in the BCDE XML format for better database integration and better data extraction.



Contact for BCDE:

mcfouzer@eecs.umich.edu

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
3600 Green Court, Suite 700, Ann Arbor, Michigan 48105-1570

Phone: 734-615-5914
Fax: 734-998-8571
http://www.ncibi.org/

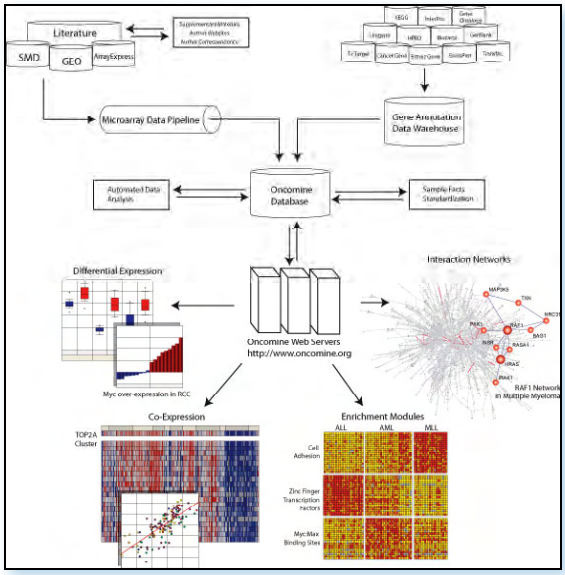
NCIBI Resources

NATIONAL CENTER FOR INTEGRATIVE BIOMEDICAL INFORMATICS

Oncomine, Cancer Profiling Database

A resource for examining gene expression in cancer

Access to this Tool: **http://www.oncomine.org**



Oncomine is a resource for examining gene expression in cancer. The goal of the project is to collect, standardize, analyze, and deliver published cancer gene expression data to the research community. Probe the expression of a gene across thousands of cancer samples or explore genes, processes, and pathways deregulated in a particular type of cancer. Oncomine pre-computes cancer profiles, clusters, and gene set modules so you can focus on discovery. Oncomine is targeted to cancer biologists who may have little bioinformatics expertise with the goal of making publicly available tumor gene expression datasets more accessible. Oncomine contains data from nearly 100 human tumor microarray studies, 18 different tumor types, and 75 million microarray measurements.

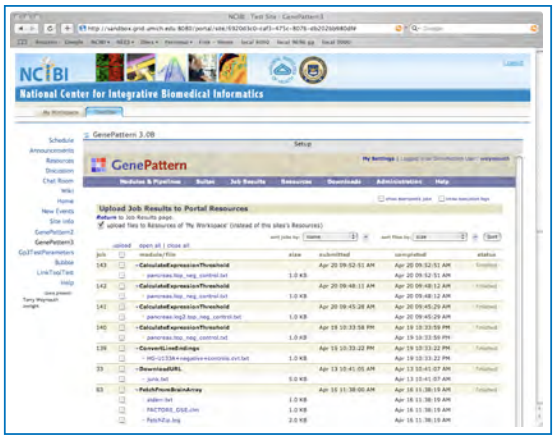
Contact for Oncomine:
oncomine@umich.edu

GenePattern

A platform for integrative genomics

Access to this Tool: **http://www.broad.mit.edu/cancer/software/genepattern/index.html**

GenePattern is an integrative computational environment that supports the analysis of data including gene expression, genomic variation (SNP), and proteomics. GenePattern's interface serves a wide community of scientists and provides an excellent environment for analytical work in computational biology, as it can easily be modified and expanded to include novel algorithms, methodologies, or visualization tools that are developed. Reproducible analytic pipelines can be easily developed, optimized, and shared across multiple collaborating organizations using the GenePattern infrastructure; A web-services architecture. The GenePattern package currently includes over 90 analysis and visualization modules and is freely available on the Internet. GenePattern was first released in 2004 and currently has nearly 5000 registered users. Support for the genomic analysis community includes live workshops an online users' forum, and other online resources. GenePattern collaborations include the NCIBI and MAGNet National Centers for Biomedical Computing, as well as the Cancer Biomedical Informatics Grid (caBIG).



Contact for GenePattern:
gp-help@broad.mit.edu

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
3600 Green Court, Suite 700, Ann Arbor, Michigan 48105-1570
Phone: 734-615-5914
Fax: 734-998-8571
http://www.ncibi.org/

NCIBI Resources

NATIONAL CENTER FOR INTEGRATIVE BIOMEDICAL INFORMATICS

MarkerInfoFinder

Genetic marker search engine for medline

Access to this Tool: **http://brainarray.mbni.med.umich.edu/brainarray/datamining/MarkerInfoFinder/**

Genome-wide high density SNP association studies are expected to identify various SNP alleles associated with different complex disorders. Understanding the biological significance of these SNP alleles in the context of existing literature is a major challenge since existing search engines are not designed to search literature for SNPs or other genetic markers. The literature mining of gene and protein functions has received significant attention and effort while similar work on genetic markers and their related diseases is still in its infancy. Our goal is to develop a web-based tool that facilitates the mining of Medline literature related to genetic studies and gene/protein function studies. Our solution consists of four main function modules for (1) identification of different types of genetic markers or genetic variations in Medline records (2) distinguishing positive versus negative linkage or association between genetic markers and diseases (3) integrating marker genomic location data from different databases to enable the retrieval of Medline records related to markers in the same linkage disequilibrium region (4) and a web interface called MarkerInfoFinder to search, display, sort and download Medline citation results. Tests using published data suggest MarkerInfoFinder can significantly increase the efficiency of finding genetic disorders and their underlying molecular mechanisms. The functions we developed will also be used to build a knowledge base for genetic markers and diseases.

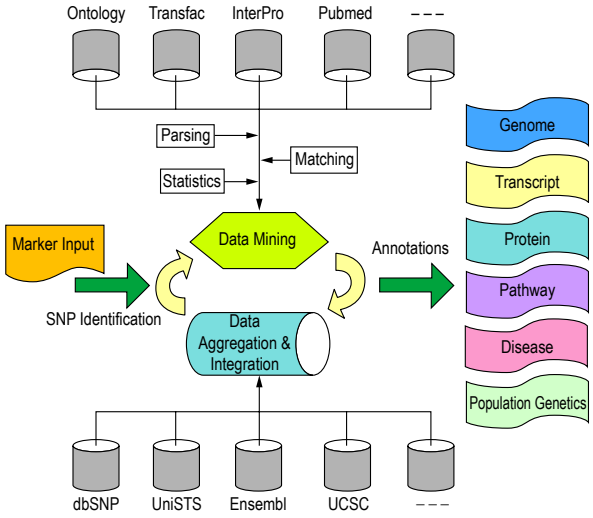
Contact for MarkerInfoFinder:
wxuan@umich.edu

SNP Functional Portal

A web database for exploring the functional implications of SNP alles

Access to this Tool: **http://brainarray.mbni.med.umich.edu/Brainarray/Database/SearchSNP/snpfunc.aspx**

Finding the potential functional significance of SNPs is a major bottleneck in understanding genome-wide SNP scanning results, as the related functional data are distributed across many different databases. The SNP Function Portal is designed to be a clearing house for all public domain SNP functional annotation data, as well as in-house functional annotations derived from different data sources. It currently contains SNP functional annotations in six major categories including genomic elements, transcription regulation, protein function, pathway, disease and population genetic. Besides extensive SNP functional annotations, the SNP Function Portal includes a powerful search engine that accepts different types of genetic markers as input and identifies all genetically related SNPs based on the HapMap Phase II data as well as the relationship among genetic markers and genes, and it greatly facilitates knowledge discovery in genome-wise SNP scanning experiments.



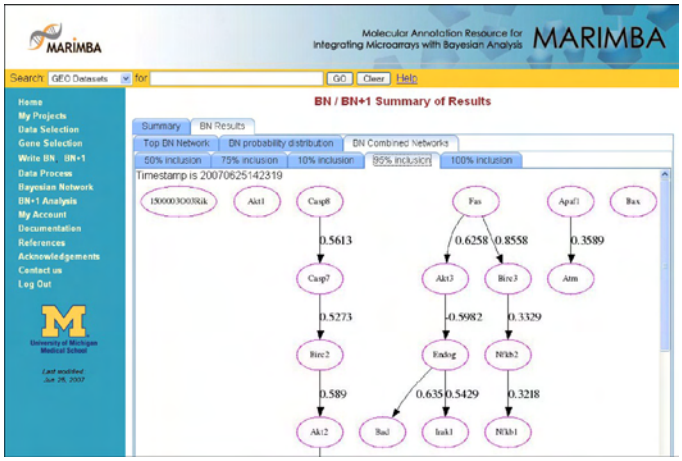
Contact for SNP Funtional Portal:
mengf@umich.edu

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
3600 Green Court, Suite 700, Ann Arbor, Michigan 48105-1570
Phone: 734-615-5914
Fax: 734-998-8571
http://www.ncibi.org/

MARIMBA: Molecular Annotation Resource for Integrating Microarray with Bayesian Analysis

Access to this Tool: **http://marimba.hegroup.org**

MARIMBA is a web-based environment for modeling biological pathways using Bayesian networks. MARIMBA integrates gene expression data and annotation from publicly available databases (e.g. NCBI GEO and AfCS) or user-defined data with existing pathway knowledge (e.g. KEGG Pathways). MARIMBA provides a user-friendly graphical interface to simplify dataset selection, gene/probeset selection, observational file processing, settings selection, Bayesian network execution, and results visualization. MARIMBA also offers automated data processing tools including fold change and clustering. MARIMBA allows users to store and update their own data and modeling results. New algorithms have been developed in MARIMBA to address fundamental pathway-related questions, including a pathway augmentation approach.

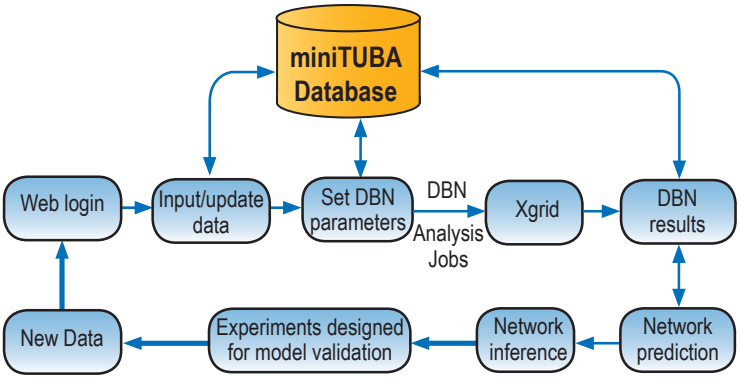


Contact for MARIMBA:
pwoolf@umich.edu, or yongqunh@umich.edu
marimba-tech@umich.edu

miniTUBA: Medical Inference by Network Integration of Temporal Data Using Bayesian Network

Access to this Tool: **http://www.minituba.org**

MiniTUBA is a web-based modeling system that allows researchers to perform complex medical/clinical inference and prediction using dynamic Bayesian network (DBN) analysis with temporal datasets. The software allows users to log in, upload and manipulate data, choose different analysis parameters (e.g., Markov lags and prior topology), execute DBN analysis, and continuously update their data and refine their results. miniTUBA can also make temporal predictions to suggest interventions based on an automated learning process pipeline using all data provided. Preliminary tests using synthetic data and laboratory research data indicate that miniTUBA accurately identifies regulatory network structures from temporal data.



Contact for miniTUBA:
pwoolf@umich.edu, or yongqunh@umich.edu
minituba-tech@umich.edu

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
3600 Green Court, Suite 700, Ann Arbor, Michigan 48105-1570
Phone: 734-615-5914
Fax: 734-998-8571
http://www.ncibi.org/

NCIBI Resources

NATIONAL CENTER FOR INTEGRATIVE BIOMEDICAL INFORMATICS

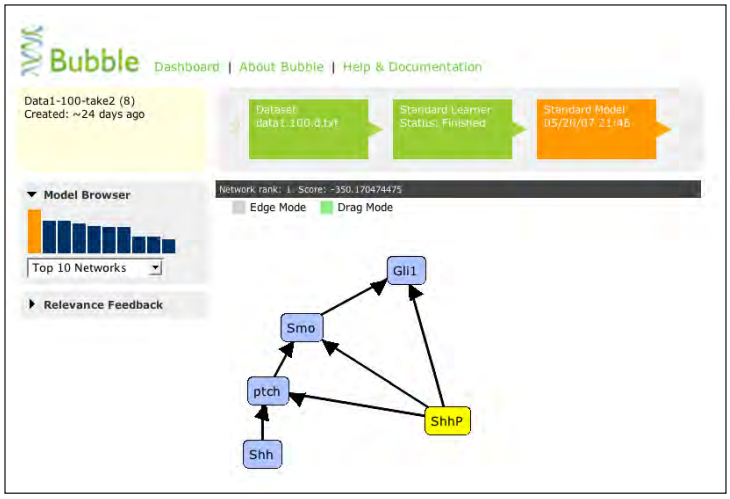
Web-Based Bayesian Network Analysis for Systems Biology (Bubble)

An interactive environment for relevance feedback for constructing Bayesian networks

Access to this Tool: **http://bubble.engin.umich.edu**

Bubble is a web-based resource for performing several Bayesian Network learning tasks on biological data. Using a simple interface, it allows users to reconstruct biological networks to identify causal relationships. Bubble features a unique entropy-reduction method to iteratively improve the learned models by selectively using lateral data from high-throughput experiments and assertions derived from literature.

Bubble will serve as a common interface for several Bayesian applications developed in our lab: a network reconstruction algorithm, a biomarker model learner, a pathway fitting application and more. Bubble's integration with the NCIBI portal allows it to be easily used in larger analysis workflows and it's use of Apple's Xgrid technology means that users can perform computationally intensive analysis from a simple web-based interface.



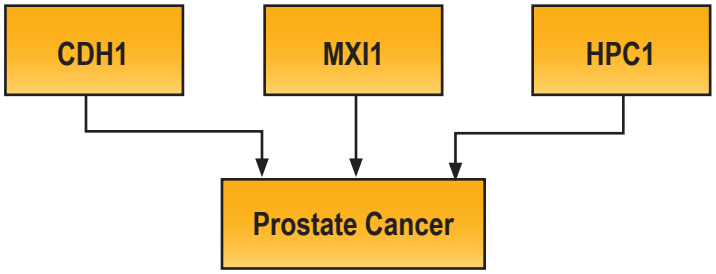
Contact for Bubble:
pwoolf@umich.edu

Markit

Biomarker discovery using Banjo

For information regarding this tool please email: **pwoolf@umich.edu**

Markit is a software tool for identifying biomarkers, such as molecular entities or clinical variables, for the purpose of predicting a clinical outcome. The software specializes in finding biomarkers which may have a nonlinear combinatorial relationship with the classifier, in contrast to many current methods which are limited to finding only linear relationships. Such methods may also overfit the data by using a complex linear model composed of many variables. Markit provides a model consisting of a small number of variables, which is not only easier for researchers to interpret, but measure as well. Additionally, biomarkers discovered using Markit are robust against measurement noise in the data. This allows for a high degree of accuracy in predicting clinical outcome when faced with new data.



Markit model example. An example of the type of Bayesian networks our algorithm attempts to discover. In this case, the parents are genes in our dataset, and the sole child node is the treatment or condition of interest. Each node has a set of possible states – for example, the 'prostate cancer' node could have a state for metastasis, localized tumor, and normal patient.

Contact for Markit:
pwoolf@umich.edu

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
3600 Green Court, Suite 700, Ann Arbor, Michigan 48105-1570

Phone: 734-615-5914
Fax: 734-998-8571
http://www.ncibi.org/

NCIBI Resources

NATIONAL CENTER FOR INTEGRATIVE BIOMEDICAL INFORMATICS

Clairlib: The Clair Library

Access to this Tool: **http://tangra.si.umich.edu/clair/clairlib/**

The **Clair library** is 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. Two distributions of the Clair library are available: Clairlib-core, with essential functionality and minimal dependence on external software, and Clairlib-ext, with extended functionality that may be of interest to a smaller audience.

Version 1.01 is now available for download.

Functionality:

Native, in Clairlib-core:
Tokenization, Summarization, LexRank, Biased LexRank, Document Clustering, Document Indexing, PageRank, Biased PageRank, Web Graph Analysis, Network Generation, Power Law Distribution Analysis, Network Analysis (clustering coefficient, degree distribution plotting, average shortest path, diameter, triangles, shortest path matrices, connected components), Cosine Similarity, Random Walks on Graphs, Statistics (distributions, tests), Tf, Idf, Perceptron Learning and Classification, Phrase Based Retrieval and Fuzzy OR Queries

Native, in Clairlib-ext:

Interface with Weka, a Java-based machine learning toolkit, LSI

Imported functionality into Clairlib-core:

Stemming, Sentence Segmentation, Web Page Download, Web Crawling, XML Parsing, XML Tree Building, XML Writing

Contact for Clairlib:
Computational Linguistics and Information Retrieval at the University of Michigan
http://tangra.si.umich.edu/clair/

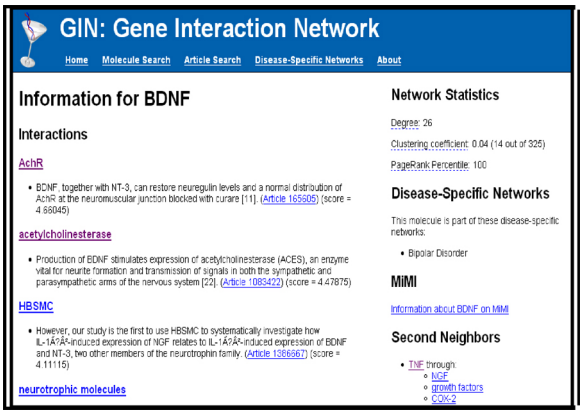
GIN: Gene Interation Network

A System for browsing articles and molecule interaction information

Access to this Tool: **http://belobog.si.umich.edu:8080/gin/**

GIN (Gene Interaction Network) is a system for browsing articles and molecule interaction information. It uses automated text mining methods (such as dependency parsing and machine learning) to extract relevant information (such as protein interactions) from text and computes statistics for the extracted interaction networks.

GIN uses articles from PubMed Central Open Access. It currently has access to 48,245 articles and 43,956 citations between these articles. A subset of the molecules in the database have been marked as disease-specific for the driving biological problems bipolar disorder, prostate cancer, type 1 diabetes, and type 2 diabetes.



Contact for GIN:
Dragomir R. Radev radev@umich.edu
Computational Linguistics and Information Retrieval

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
3600 Green Court, Suite 700, Ann Arbor, Michigan 48105-1570

Phone: 734-615-5914
Fax: 734-998-8571
http://www.ncibi.org/

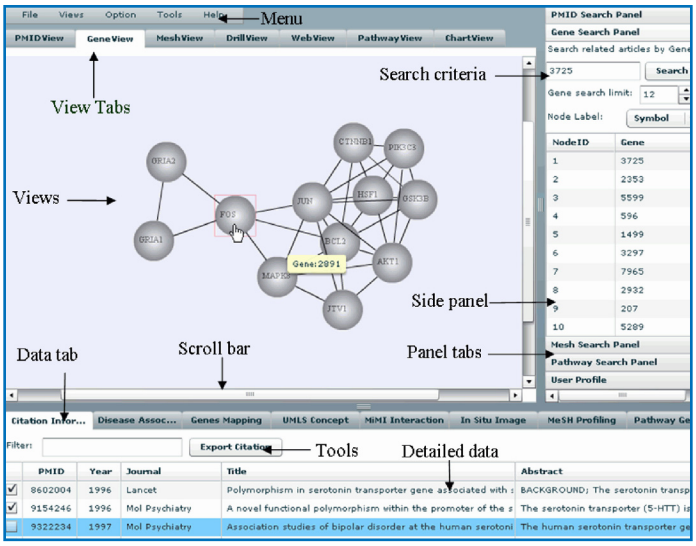
NCIBI Resources

NATIONAL CENTER FOR INTEGRATIVE BIOMEDICAL INFORMATICS

PubViz An Interactive Search Interface for Medline

Access to this Tool: **http://www.pubviz.org**

Searching the Medline database is almost a daily necessity for many biomedical researchers. However, current Medline search solutions are mainly designed for the quick retrieval of a small set of most relevant documents. Because of this search model, they are not suitable for the large-scale exploration of literature and the underlying biomedical conceptual relationships, which are common tasks in the age of high throughput experimental data analysis and cross-discipline research. We developed a new Medline exploration approach by incorporating interactive visualization together with powerful grouping, summary, sorting and active external content retrieval functions. Our solution, PubViz, is based on the FLEX platform designed for interactive web applications and its prototype.



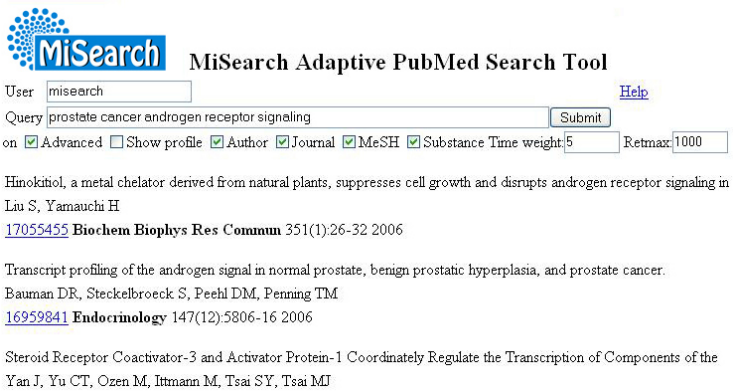
Contact for PubViz:
wxuan@umich.edu and http://www.pubviz.org/

MiSearch

Adaptive PubMed search tool

Access to this Tool: **http://misearch.ncibi.org**

MiSearch is an adaptive literature search tool designed to facilitate the identification of relevant citations in the PubMed database. MiSearch captures user behavior during searching and browsing citations and builds a statistical profile describing the articles a user has selected. It then uses this profile to rank the results of future searches, placing those articles most likely to be selected at the top of the list. MiSearch uses the NCBI Entrez Eutils interface to search a query against the PubMed database and ranks the results using a statistical profile ranking those articles that a user is most likely to have selected in past searches. The profile incorporates four features: Medical Subject Headings (MeSH), substance names, journals, and the names of the authors. In advanced search mode, the user is able to specify which of these features to include and how heavily to weight the date of publication.



Contact for MiSearch:
misearch-help@umich.edu

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
3600 Green Court, Suite 700, Ann Arbor, Michigan 48105-1570
Phone: 734-615-5914
Fax: 734-998-8571
http://www.ncibi.org/