



## Integrating Tools and Data: NCIBI Suite of Tools

National Center for  
Integrative Biomedical Informatics  
([www.ncibi.org](http://www.ncibi.org))



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## National Centers for Biomedical Computing (NCBCs)

- The National Centers for Biomedical Computing (NCBC) are cooperative agreement awards that are funded under the NIH Roadmap for Bioinformatics and Computational Biology
- The Centers are intended to be the core of the networked national effort to build the computational infrastructure for biomedical computing in the nation, the National Program of Excellence in Biomedical Computing (NPEBC)
- There are seven funded Centers that cover systems biology, image processing, biophysical modeling, biomedical ontologies, information integration, and tools for gene-phenotype and disease analysis
- The centers will create innovative software programs and other tools that enable the biomedical community to integrate, analyze, model, simulate, and share data on human health and disease

o <http://www.ncbcs.org/> o <http://www.bisti.nih.gov/ncbc/>



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## NCIBI

[www.ncibi.org](http://www.ncibi.org)

*The mission of the National Center for Integrative Biomedical Informatics (NCIBI) is to create targeted knowledge environments for molecular biomedical research that help guide experiments and enable new insights from the analysis of complex diseases. NCIBI develops efficient software tools, data integration methods, and systems modeling environments. The resulting NCIBI "Suite of Tools and Data" facilitates rapid construction of context-appropriate molecular biology information schemas from experimental data*



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## Today's Demonstration

- Use a biological case-study to demonstrate the utility and integration of tools developed by NCIBI
- Featured tools:
  - MiMI (Michigan Molecular Interactions)
  - MiSearch (Adaptive Literature Search Tool)
  - GIN (NLP literature summarization and centroiding)
  - Gene2MeSH (Gene / MeSH Term matrix)
  - NetworkBrowser (Flex-based network visualization tool)
  - BioSearch-2D (Heat-map viewer for literature (gene/MeSH) interactions)
  - MiMI Plugin for Cytoscape (API and visualization for the MiMI Data and linkage to other Cytoscape Plugin tools)
  - SAGA (subgraph approximate matching tools for network similarity)
  - BioNLP (parsed and gene-tagged version of PubMed and PMCOA)



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## MiMI

[mimi.ncibi.org](http://mimi.ncibi.org)

- MiMI is the Michigan Molecular Interactions database
- It is a composite database of many existing protein-protein interaction databases and protein-gene information as well as literature based protein-protein inferred interactions (in final stages)
- Maintains provenance of the source and nature of the data included



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## MiSearch

[misecond.ncibi.org](http://misecond.ncibi.org)

- MiSearch is a literature search tool that works with NCBI Entrez, and you can use it to rapidly search PubMed citations and see them displayed automatically in order of those most relevant to your research interests
- This relevance ranking is based on a profile of your areas of interest that MiSearch stores from past searches you conduct with MiSearch. MiSearch saves information on citations that you selectively view from the results of a query, and it uses this information to build a statistical profile describing your choices. To do so, MiSearch uses a classification algorithm based on MeSH term, substance names, journal, and author names associated with citations
- Use this tool to get the same results as a PubMed search but ranked by a “personalized” relevance to your topic



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## GIN

[gin.ncibi.org](http://gin.ncibi.org)

- GIN (Gene Interaction Network) is a system for browsing articles and molecular interaction information. GIN uses automated methods (such as dependency parsing) to mine the text for relevant information, for example protein interactions. It also computes statistics for the interaction network. The user can see summary sentences highlighted. The user can browse molecules to view their interactions, neighborhood, and other network statistics
- GIN uses text mining and network analysis methods to predict gene-disease associations. Disease-specific interaction networks are built by starting with initial lists of seed disease genes that are known to be related to a disease. All the interactions among the seed genes and the genes that interact with them are extracted automatically from the literature. Centrality metrics are used to rank the genes in the constructed disease-specific networks. The hypothesis is that the central genes in these networks (inferred disease genes) are likely to be related to the diseases



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## Gene2MeSH

[gene2mesh.ncibi.org](http://gene2mesh.ncibi.org)

- 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. Gene2MeSH also provides relevant links to protein interactions in MiMI as well as reference links to Entrez, the MeSH browser, and PubMed
- *Metab2MeSH is a related tool linking MeSH headings with biological substances (Metabolites) in a similar fashion (still in prototype) (<http://metab2mesh.ncibi.org>)*



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## NetworkBrowser

- Gene/NetworkBrowser is a light-weight web-based tool for visualizing networks. It is based on the Flex language and is linked to our databases (e.g MiMI and BioNLP) and to existing biological annotation such as GO, MeSH headings and KEGG
- It is a portable and easily implemented interface that we are prototyping for visualization
- This tool is currently only called from within other tools

– Please email for information: [cavalcol@umich.edu](mailto:cavalcol@umich.edu)



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## BioSearch-2D

[biosearch2d.ncibi.org](http://biosearch2d.ncibi.org)

- BioSearch-2D renders the contents of large biomedical document collections into a single, dynamic map
- With this engine, users can generate a summary map of genes vs. ontology topics which match that provided by expert human review articles
- In addition, BioSearch-2D provides users a context-specific functional annotation system for high-throughput experiment gene signatures and ad-hoc gene lists



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## MiMI Plugin for Cytoscape

[mimiplugin.ncibi.org](http://mimiplugin.ncibi.org)

- The Cytoscape MiMI Plugin is an open source interactive visualization tool that you can use for analyzing protein interactions and their biological effects
- The Cytoscape MiMI Plugin couples Cytoscape, a widely used software tool for analyzing biomolecular networks, with the MiMI database, a database that uses an intelligent deep-merging approach to integrate data from multiple well-known protein interaction databases
  - The MiMI database has data on 119,880 molecules, 330,153 interactions, and 579 complexes
- By querying the MiMI database through Cytoscape you can access the integrated molecular data assembled in MiMI and retrieve interactive graphics that display protein interactions and details on related attributes and biological concepts



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## SAGA

[saga.ncibi.org](http://saga.ncibi.org)

- SAGA (Substructure Index-based Approximate Graph Alignment) is a tool for querying a biological graph database to retrieve matches between subgraphs of molecular interactions that scientists select and biological networks
- SAGA implements an efficient approximate subgraph matching algorithm that can be used for a variety of biological graph matching problems. One application is: pathway matching, in which SAGA is used to compare pathways in KEGG and Reactome
- You can also use SAGA to find matches in literature databases that have been parsed into semantic graphs. In this use of SAGA, portions of PubMed have been parsed into graphs that have nodes representing gene names



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting



## BioNLP

- PubMed and PMCOA are freely usable resources. NCIBI has parsed the existing text literature according to several existing and some newly developed algorithms. A well developed schema makes SQL access to this data straightforward
- Genes are efficiently tagged for presence in sentences using a combination of statistical and dictionary-based taggers
- 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
- Currently BioNLP is only available as a foundation of our tools via SQL. Other user access may be possible
  - *Please email for access and details: [cavalcol@umich.edu](mailto:cavalcol@umich.edu)*



National Center for Integrative Biomedical Informatics - 2008 NCBC All Hands Meeting

