

# **Post-Genomic Challenges For the Open Source Bioinformatics Community:**

**Systems Biology as a Target Source**

**Dan Kilburn, PhD**

**Beyond Genomics, Inc.**

# Bioinformatics in 2002

- Transition from a gene-centric to systems-oriented approach to knowledge discovery
- Introduction to Systems Biology
- Overview of approaches and methods in need of “Open Source Intervention”

# The Hegelian Dialectic and the Race to Sequence the Human Genome

- **Thesis:** Gene Myers, Craig Venter and the Apocryphal Cocktail Napkin
- **Antithesis:** Ensembl and the OpenBio Movement
- **Synthesis:** Demise of “Genomes For Sale”
- *What’s the new thesis?*

## The Dialectic, Part II

- Pre-1998 Bioinformaticists often worked in labs headed by biologists. Problem: it's always difficult to work for someone who doesn't understand what you do
- 1998-2001 The Genome Utopia: Genomic data was available over the web at a rate of ~100 megabases a week, freeing us from the tyranny of biological rule and worrying about the messiness of where/how data was generated
- Post 2001 The Dream is over: now we have to work with biologists and raw data again



# Ensembl: A Gene-Centric Genome Viewer

Ensembl Human Genome Server (ContigView) - Microsoft Internet Explorer

Address: [http://www.ensembl.org/Homo\\_sapiens/contigview?chr=1&vc\\_start=1&vc\\_end=1000000&w=0&x=0](http://www.ensembl.org/Homo_sapiens/contigview?chr=1&vc_start=1&vc_end=1000000&w=0&x=0)

Find Sequence:   [e.g. [U34879](#), [AP000869](#)]

### Chromosome 1

Chr 1

### Overview

Chr 1 band: 0 bp to 1,000,000 bp

Gene Legend: ENSEMBL\_PREFERRED GENES (KNOWN), ENSEMBL\_PREFERRED GENES (NOVEL), ENSEMBL\_COHORTED GENES

### Detailed View

Jump to chr:  bp  to

Zoom:

Features: DAS Sources, Decorations, Export, Jump to Image size

99%

9:58 AM

# Jim Kent's BLAT Search Engine

The screenshot shows the BLAT Search Engine web page. At the top, the browser title is "BLAT Search - Microsoft Internet Explorer". The address bar shows the URL: <http://genomes.uscc.edu/cgi-bin/blast?command=blast&start=1012>. The page has a navigation menu with "File", "Edit", "View", "Favorites", "Tools", and "Help". Below the menu, there are buttons for "Back", "Forward", "Search", "History", "Print", "Page Info", "Stop", and "Links". The main content area has a title "BLAT Search Genome". There are several dropdown menus: "Freeze:" set to "June 2002", "Query type:" set to "BLAT's guess", "Sort output:" set to "query score", and "Output type:" set to "hyperlink". A "Submit" button is located to the right of the "Output type:" dropdown. Below these menus is a text input field for a query sequence, followed by "Browse..." and "Submit File" buttons. A large text area for results is visible below the input field. The page contains several paragraphs of text explaining the BLAT search engine's capabilities and limitations.

BLAT Search Genome

Freeze:  Query type:  Sort output:  Output type:

Please paste in a query sequence to see where it is located in the genome. Multiple sequences can be searched at once if separated by a line starting with > and the sequence name.

Rather than pasting a sequence, you can choose to upload a text file containing the sequence

Upload sequence:

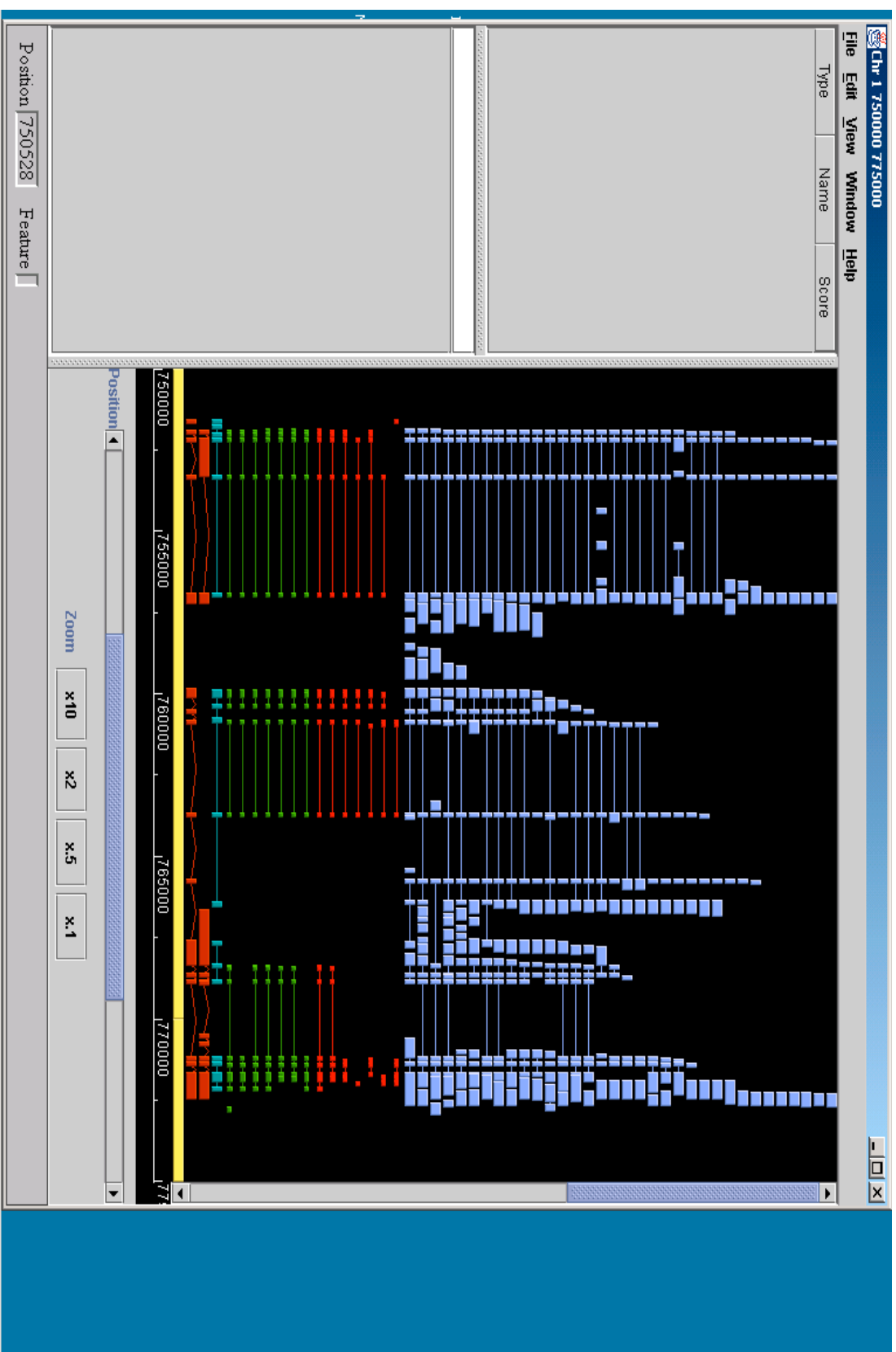
Only DNA sequences of 25,000 or less bases and protein or translated sequence of 5000 or less letters will be processed. If multiple sequences are submitted at the same time, the total limit is 50,000 bases or 12,500 letters.

BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 40 bases or more. It may miss more divergent or shorter sequence alignments. It will find perfect sequence matches of 35 bases, and sometimes find them down to 22 bases. BLAT on proteins finds sequences of 80% and greater similarity of length 20 amino acids or more. In practice DNA BLAT works well on primates, and protein blat on land vertebrates

BLAT is not BLAST. DNA BLAT works by keeping an index of the entire genome in memory. The index consists of all non-overlapping 11-mers except for those heavily involved in repeats. The index takes up a bit less than a gigabyte of RAM. The genome itself is not kept in memory, allowing BLAT to deliver high performance on a reasonably priced Linux box. The index is used to find areas of probable homology which are then loaded into memory for a detailed alignment. Protein BLAT works in a similar manner, except with 4-mers rather than 11-mers. The protein index takes a little more than 2 gigabytes

BLAT was written by [Jim Kent](#). Like most of Jim's software interactive use on this web server is free to all. Sources and executables to run batch jobs on your own server are available free for academic, personal, and non-profit purposes. Non-exclusive commercial licenses are also available. Contact Jim for details.

# The Apollo Genome Browser



# Genome Browsers: Challenges to Overcome / A Wish List

**Static:** underlying data is updated sporadically

**Noninteractive I:** users have little or no capacity to integrate proprietary data

**Noninteractive II:** browsers have little or no capacity to respond to or record users' path through data

**Orthogonal:** browsers depict physical relationships and gene family relationships, but not biologically-relevant ones like receptor-ligand and other partner relationships, biological processes, cellular and tissue localization.

# DAS: Integrating Distributed Data

**Resources**

- [Biodas Home](#)
- [DAS Overview](#)
- [DAS/1 Specification](#)
- [DAS/2 RECS](#)

**About Biodas**

Welcome to biodas.org. This site is the center of development of an Open Source system for exchanging annotations on genomic sequence data. Here is an [overview](#) of DAS written by Scott Pearson.

If you came here looking for the Dense Alignment Surface method, please refer to <http://www.spc.su.se/~miklos/>.

**About the DAS Project**

The distributed annotation system (DAS) is a client-server system in which a single client integrates information from multiple servers. It allows a single machine to gather up genome annotation information from multiple distant web sites, collate the information, and display it to the user in a single view. Little coordination is needed among the various information providers.

The original version 1 specification, written by Lincoln Stein and Robin Dowell, is the basis for a number of clients and servers. DAS/1 servers are currently running at [XombaBase](#), [Eukbase](#), [Ensembl](#), [HISK](#), and [UCSC](#).

In light of lessons learned during the DAS/1 implementation, we are planning a transition to a more flexible and powerful protocol called DAS/2. This transition will occur slowly, and DAS software will continue to support DAS/1 for some years to come. We are currently collecting RFCs (Requests for Comment) on DAS/2. To read the RFCs or contribute your own, please see [the DAS/2 RFC page](#).

Currently the DAS code base consists of:

1. The DAS/1 specification (stable)
2. The preliminary [DASXFE](#) specification, used by the Dazzle client (subject to change)
3. Complete DAS clients:
  - o [Ossodestic](#), by Robin Dowell
  - o [DasClient](#), by Matthew Pocock
  - o [DrmDAS/Commsense](#)
4. Client Libraries
  - o [BioJava](#) client library (DAS/1 & DAS/2)
  - o [Bio.DAS.perl](#) client library (DAS/1 only)
5. Servers
  - o [Dazzle Java server](#)
  - o [Lightweight DAS server \(Perl\)](#)

**Mailing List**

The DAS developer's mailing list is [das@biodas.org](mailto:das@biodas.org). To subscribe, go to [this page](#).

**Public DAS Servers**

Foo Cheung is assembling a list of publicly-accessible DAS servers. If you have a server that is not on the list and you would like it to appear there, please send him mail.

**News**

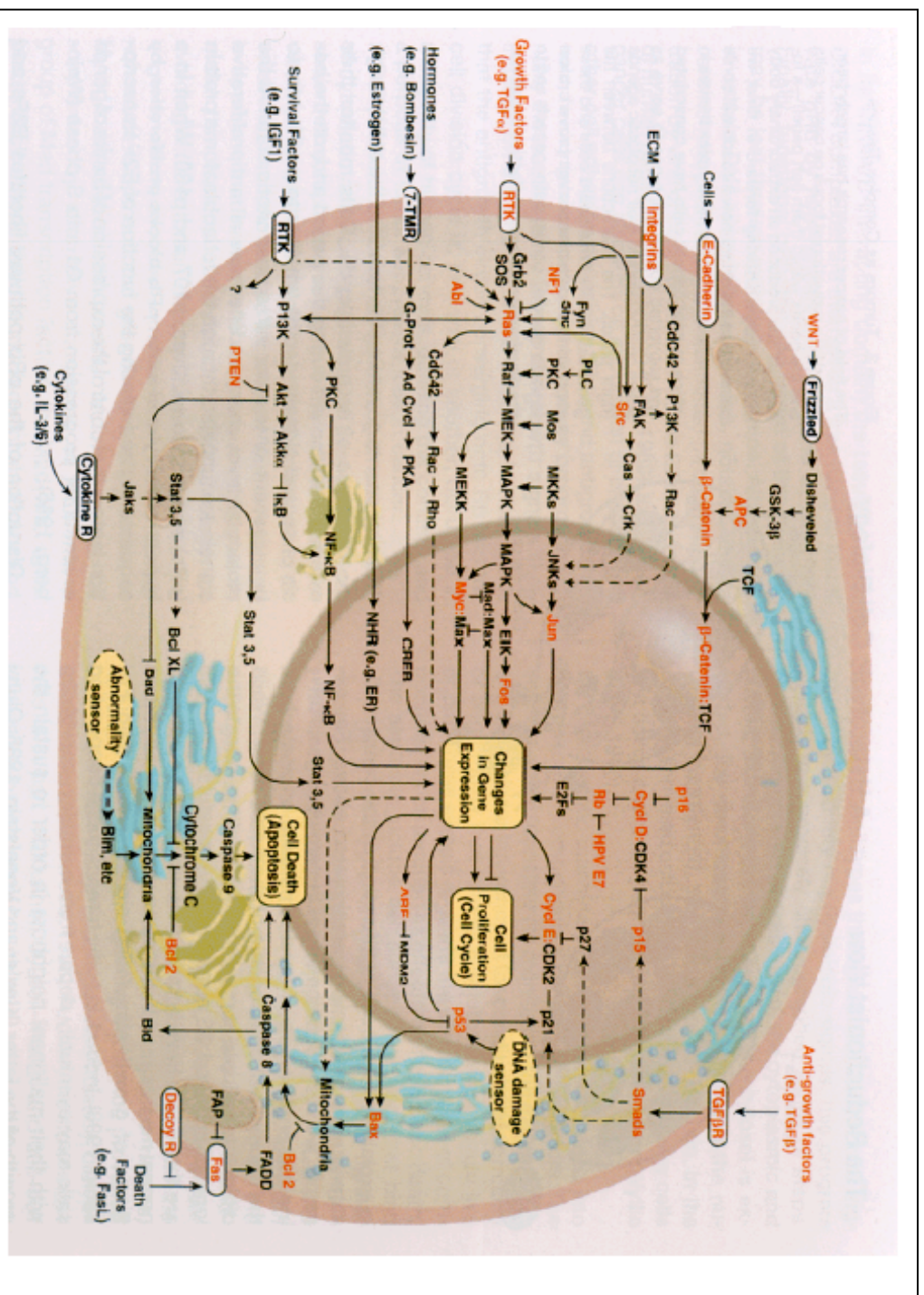
- July 23, 2002  
Version 1.07 of the [LDAS server](#) fixes a bug that caused the DSN response to include unescaped ampersands and other entities.
- April 22, 2002  
Version 1.06 of the [LDAS server](#) fixes a bug that prevented links from being generated for many features.
- April 15, 2002  
Stable version 0.20 of [Bio::Das.perl](#) client library released.  
Bio::DAS.perl client library (DAS/1 only)  
Experimental version 0.6 of [Bio::Das](#) released (parallel fetch, but no support for stylesheets).
- February 11, 2001  
Version 1.5 of the DAS specification has been released. This is an interim release designed to fix some glaring deficiencies in 1.0 spec, but backwardly compatible with earlier versions.  
Updated address for Foo Cheung's list of DAS servers.  
December 17, 2001  
Added Foo Cheung's list of public DAS servers.  
December 7, 2001  
Version 1.01 of the [LDAS server](#) fixes a bug in database authentication.  
November 26, 2001  
Foo Cheung has written a new Web-based DAS viewer. It is in early form, but the source will eventually be contributed to the DAS repository. It can be viewed at <http://www.tigr.org/tdb/DAS/dasview.html>
- November 20, 2001  
Biodas mailing list moved to [biodas.org](mailto:biodas.org)
- November 5, 2001

# Foreshadowing

- Moving Beyond the Genome Utopia
- Standards to Cope with Proliferation of Platforms and Vendors
- Lims Systems and Pipeline Management
- Data Analysis Standards
- Visualization Tools
- Working with Clusters while Avoiding Bankruptcy



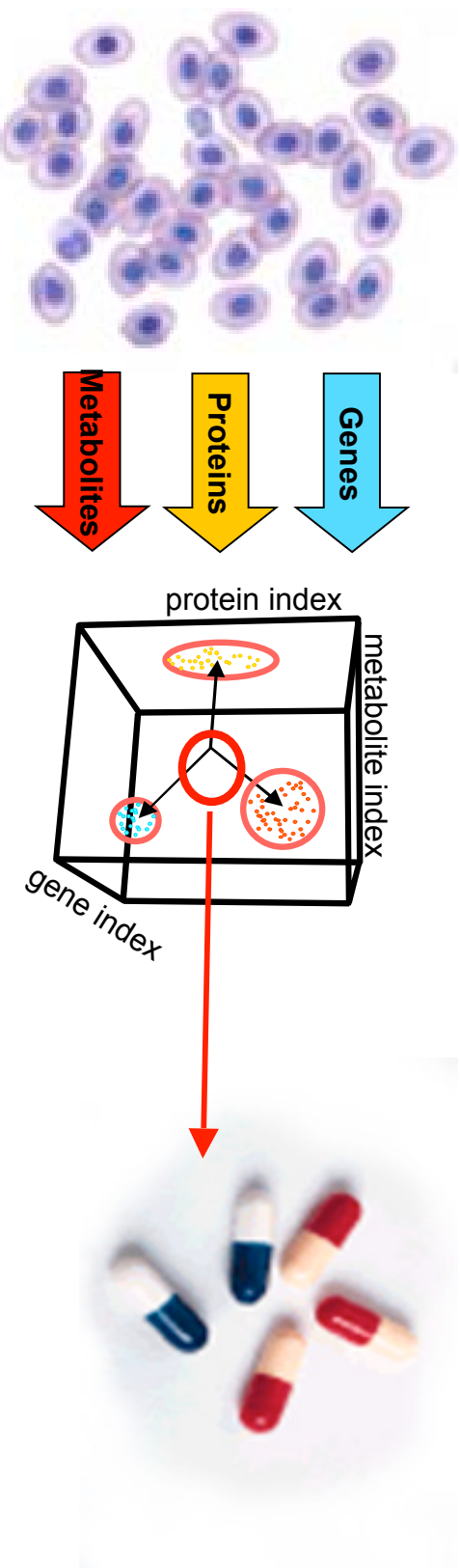
# Biological Systems



D. Hanahan and R. A. Weinberg. Cell., 100(1):57-70 Review, 2000.

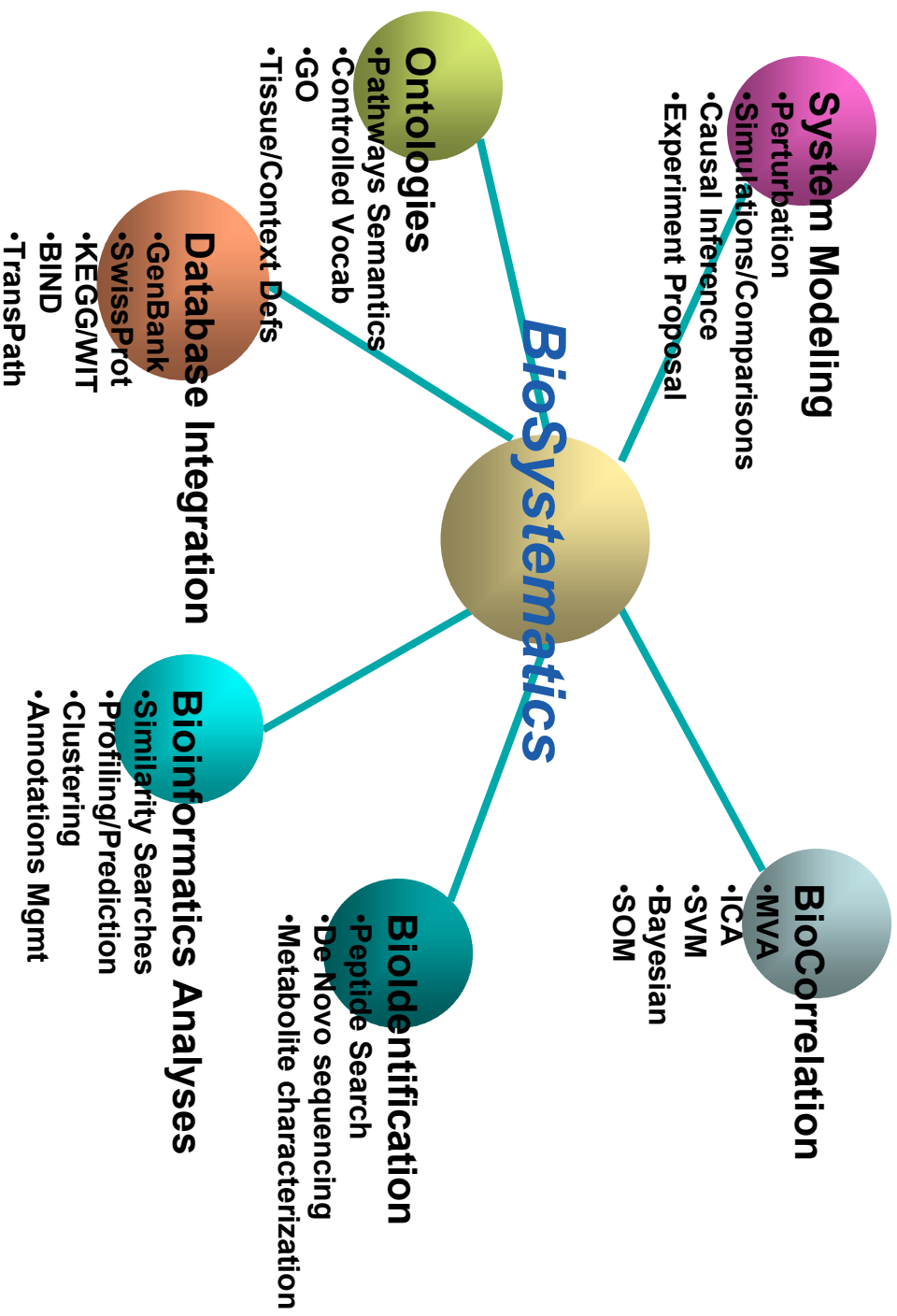
# What is Systems Biology?

- Parallel analyses of proteins, metabolites and mRNA from complex samples
- Perturbations using mutations, drugs, and experimental conditions
- Determination of molecular function and elucidation of cellular mechanisms
- Informatics tools to link gene response, protein activity and metabolite dynamics
- *BioSystematics* translates covariant sets of genes, proteins, metabolites into *biochemical interaction and target information*

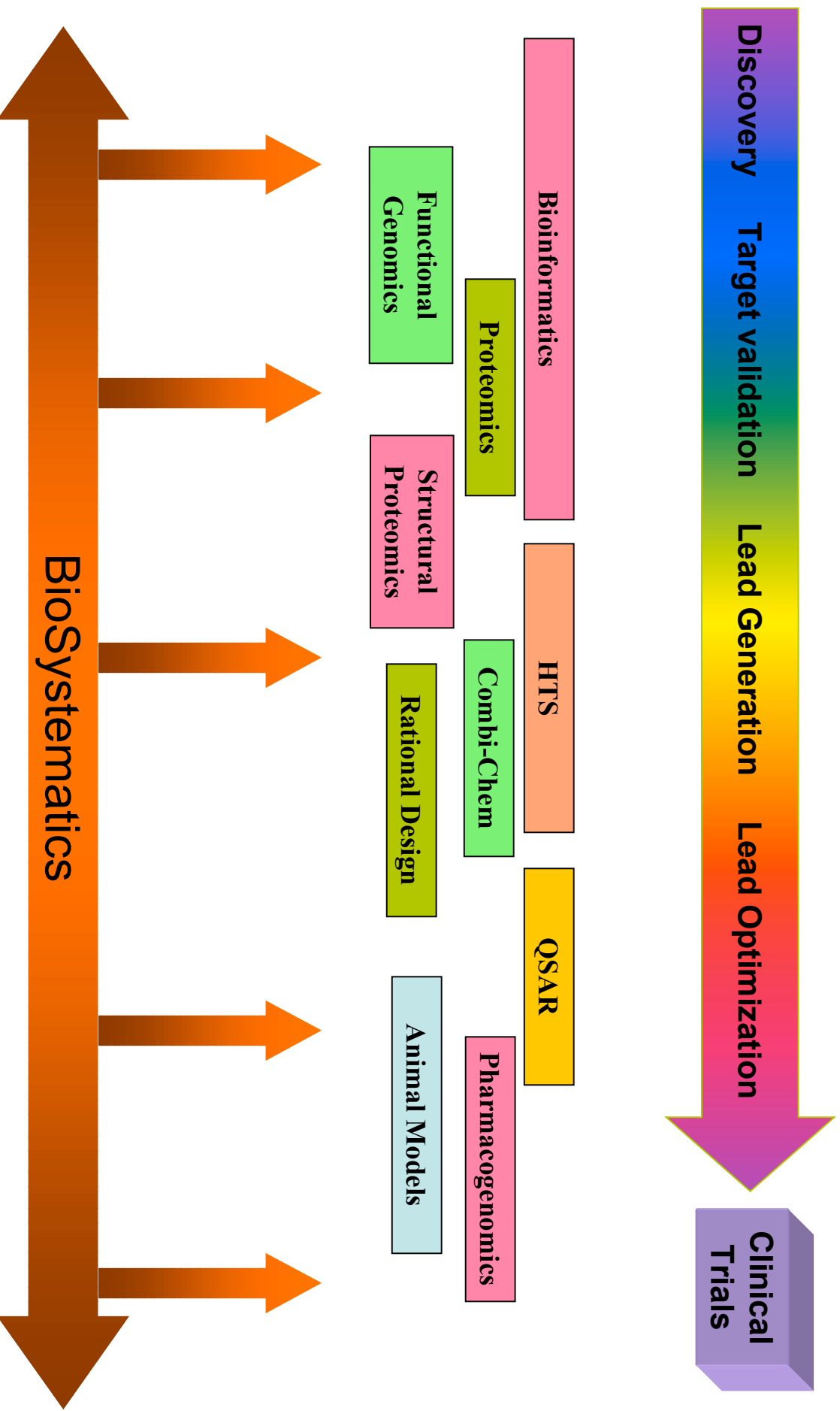




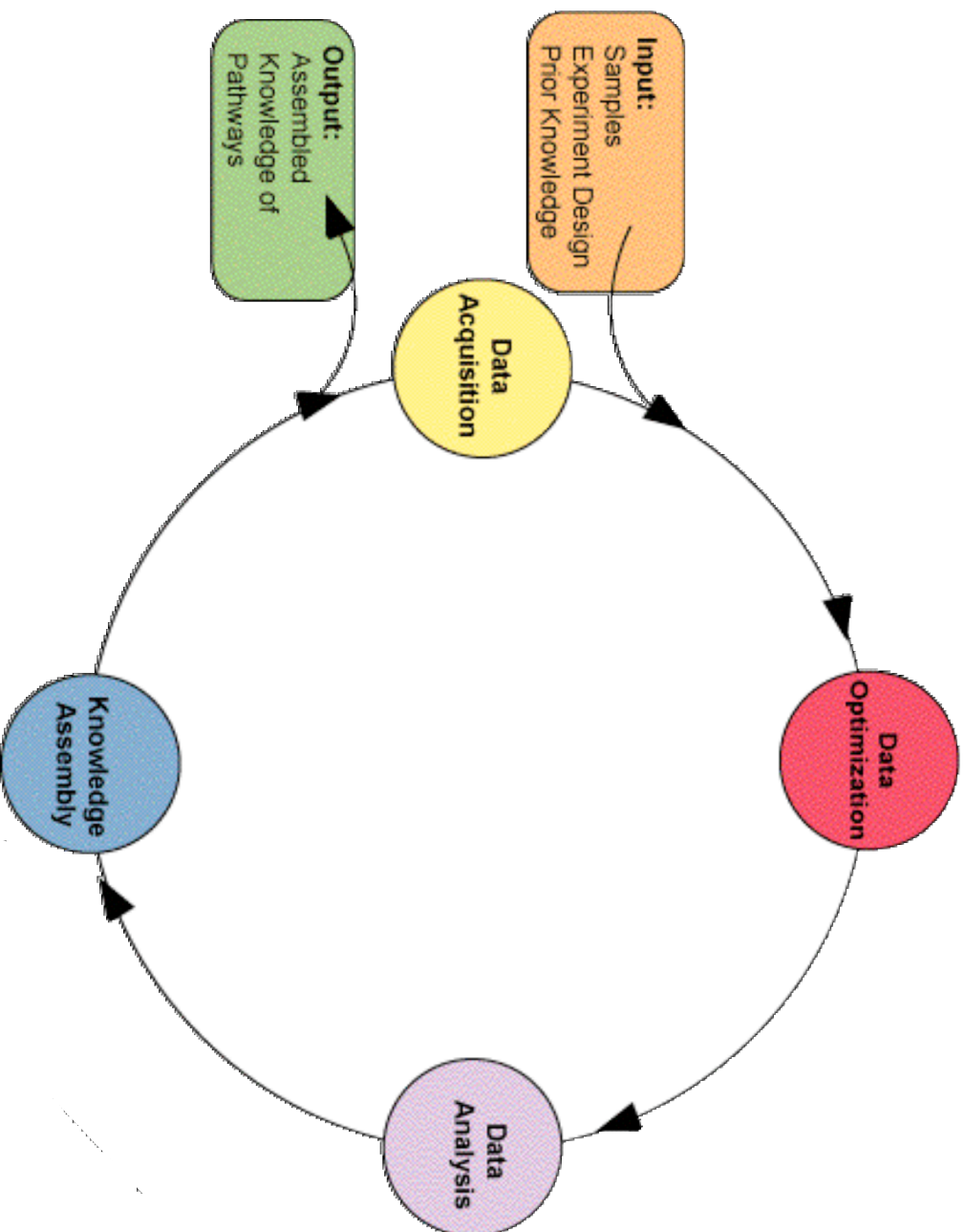
# Systems Biology + Bioinformatics =



# Role of BioSystematics in Drug Discovery



# BioSystematics and the Scientific Method

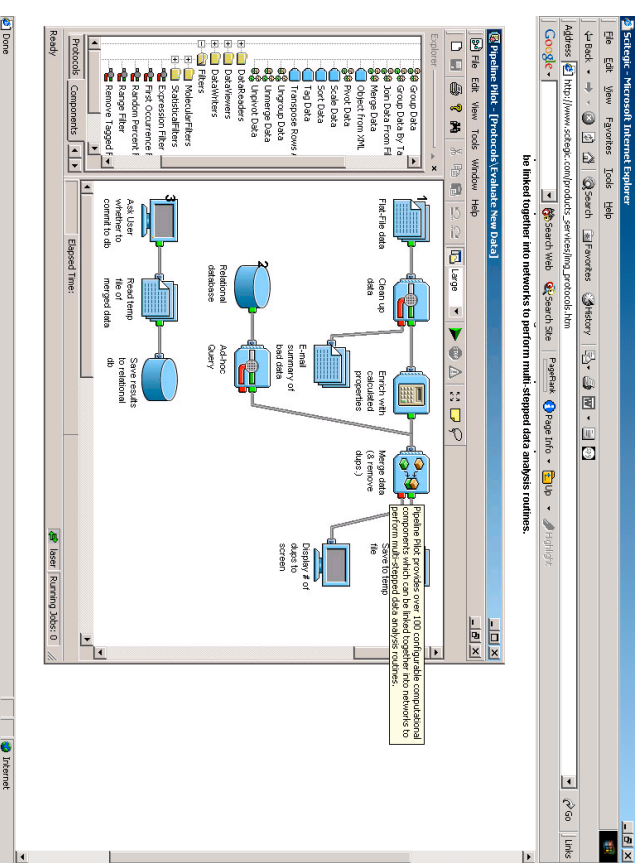


# Analysis Pipelines

- A real world example: TF identification
- Two private sector approaches:
  - Scitegic
  - Incogen
- Open Source Initiatives

# Scitegic

- What Is Data Pipelining?
  - The processing, analysis, and mining of large volumes of data through a user-defined computational protocol.
  - Data Pipelining: Data is progressively read into a protocol and analyzed or modified by each component as it flows through the network toward the output endpoints.



# IncoGen

Address: <http://www.incoGen.com/yibe/pipeline.gif>

File Edit View Favorites Tools Help

Search Target DB Utility Seals Visualization

SW NT SW AA BLASTN BLASTX BLASTP BLASTZ HMMSearch HMMScan

Parameter Value

ALGORITHM	SW
ALIGNMENT_THRESHOLD	20
COMMENT	Smith-Waterman Search
EOL	LF
EXTEND_PENALTY	2
JOB_ID	
MATRIX	BLOSUM62
MAX_ALIGNMENTS	20
NULL_CHARACTER	30
OPEN_PENALTY	
PSCORE	F1.2
QUERY_FORMAT	GAPPED
QUERY_NEIGHBORHOOD	FAST/PEARSON
QUERY_PATH	
QUERY_SEARCH	F
QUERY SET	
QUERY SPLIT	
QUERY TYPE	AA
SCALE FACTOR	F
SIGNIFICANCE	PSCORE GAPPED
TARGET_FRAMES	1
TARGET SET	swisprot, swplus
TARGET TYPE	AA
THRESHOLD	F

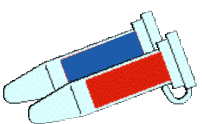
Details Parameters Notes

Kinase 1 Kinase 2

Internet

# Proteomics Analysis

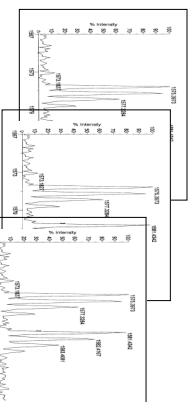
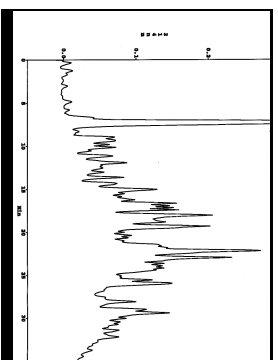
- control sample
- experimental sample



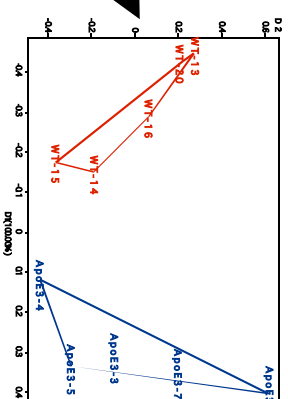
sample preparation chemistry / isotopic labeling



multi-dimensional chromatographic separation



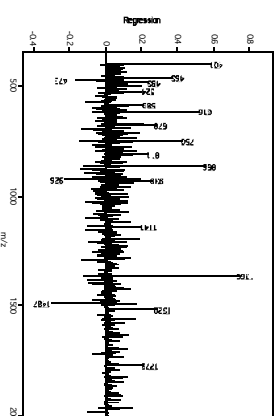
collection and mass spectrometry



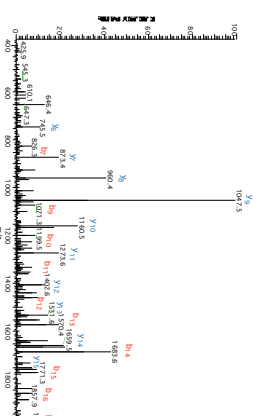
clustering of datasets



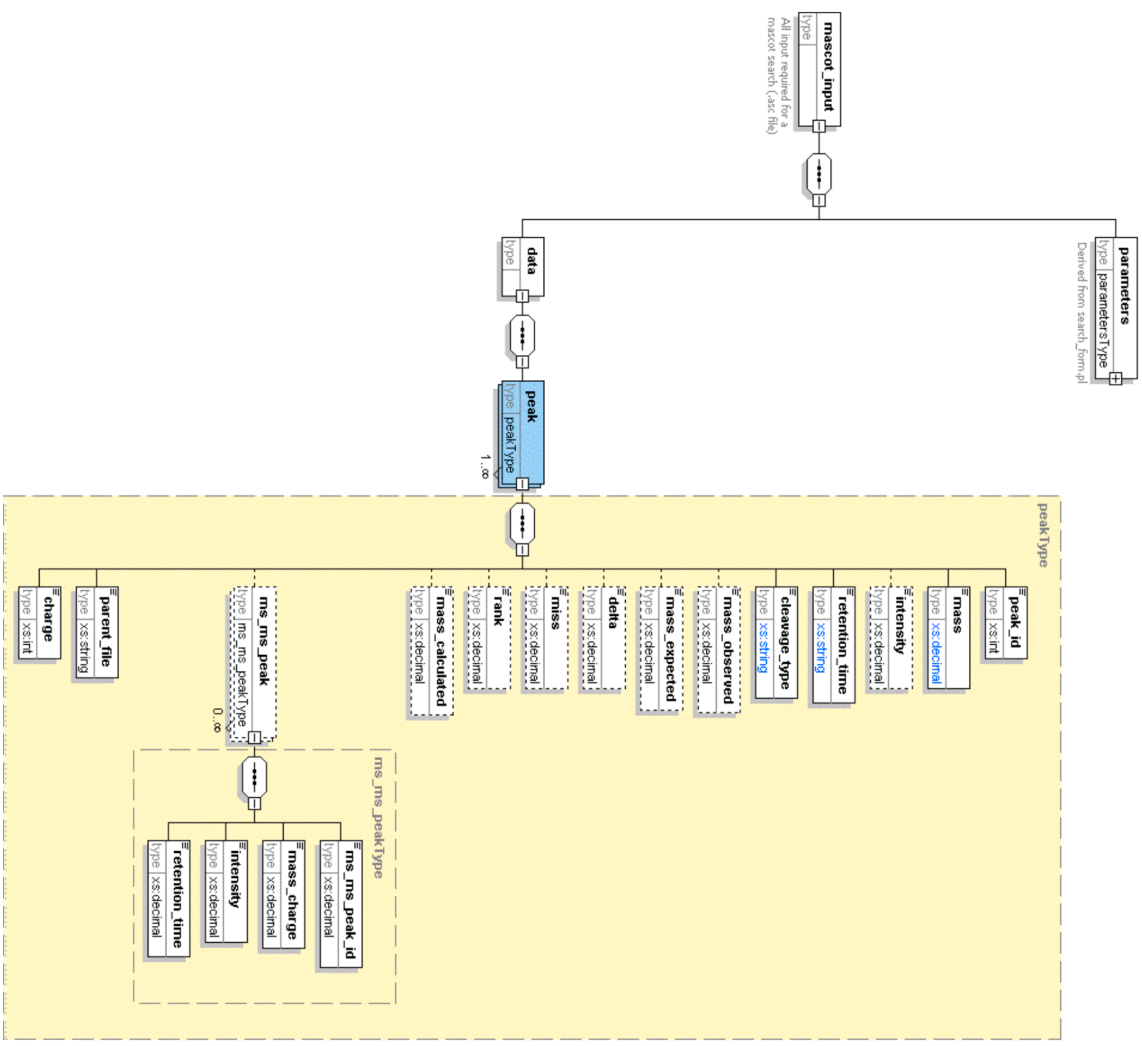
pattern recognition of significant peptide differences



MS/MS identification of relevant peptides

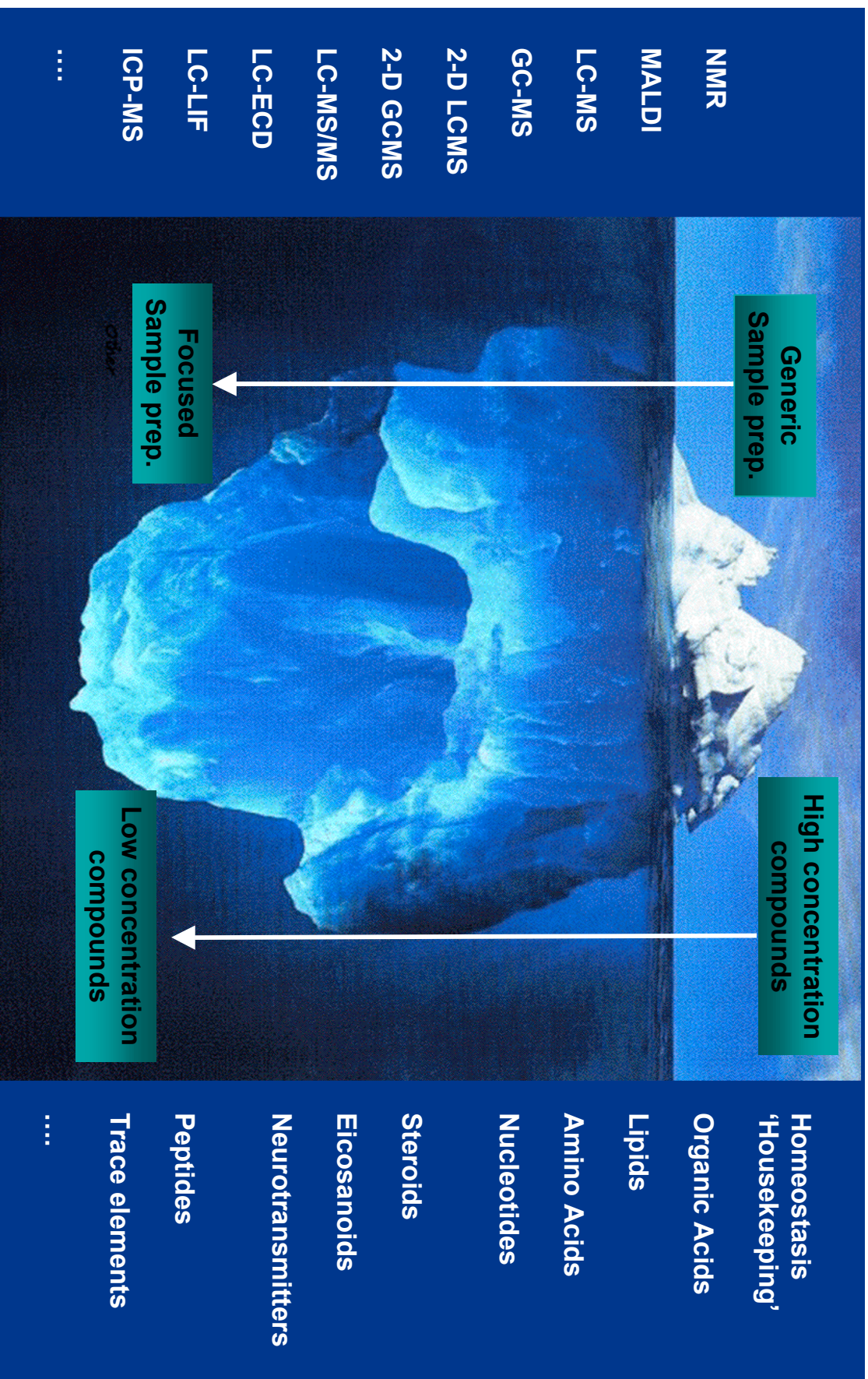


# MS Peaks and Proteins XML





# Metabolomics



# Data Acquisition: LIMS *Expt Design*

Microsoft Internet Explorer  
BG LIMS - Microsoft Internet Explorer

Address: http://lmsapp1.lifescience.com/BGlims/main.jsp

File Edit View Favorites Tools Help

Back Forward Stop Search Search Search PageRank Page Info Up Highlight

Beyond Genomics LIMS  
Dan Kilburn

LIMS System

ABC Pharma

Project	Type	Start Date	End Date	Delete
Test		05/14/2002		Delete

Beyond Genomics

Project	Type	Start Date	End Date	Delete
Mouse Platform Dev	Internal	03/19/2002		Delete

diaDexus

Project	Type	Start Date	End Date	Delete
Phase 1A Norm	Serum	04/29/2002		Delete

Elan

Project	Type	Start Date	End Date	Delete
PDAPP Mouse	external	03/19/2002		Delete

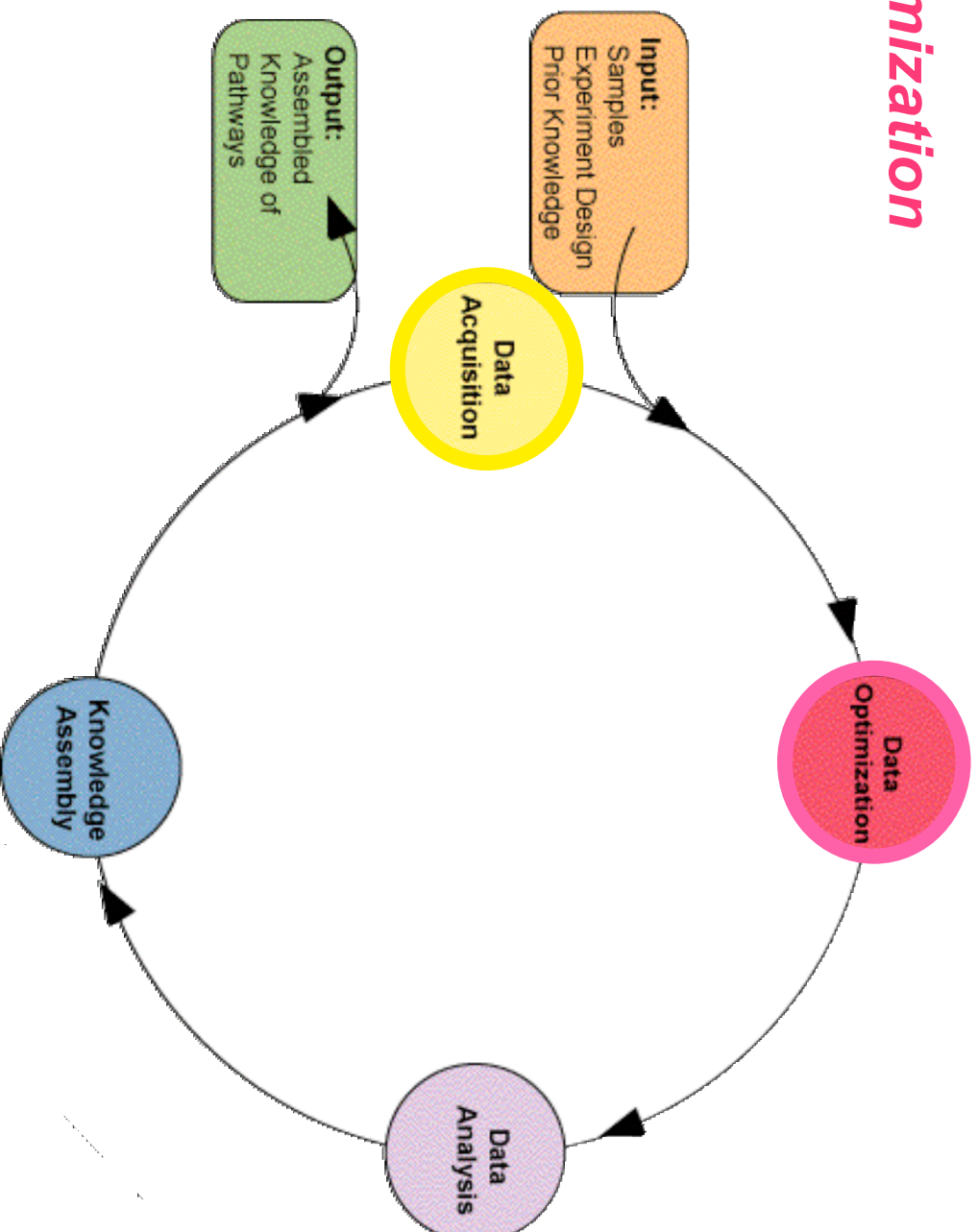
Navigation Menu:

- LIMS System
  - ABC Pharma
  - Beyond Genomics
  - diaDexus
  - Elan
  - System Manager
  - Users
  - Containers
  - Packages
  - Process Definitions
  - Protocols
  - States
  - Storage Locations

Taskbar: Start, Internet Explorer, Outlook, My Documents, 99% battery, 9:50 AM

# BioSystematics™ at Beyond Genomics:

## Data Optimization

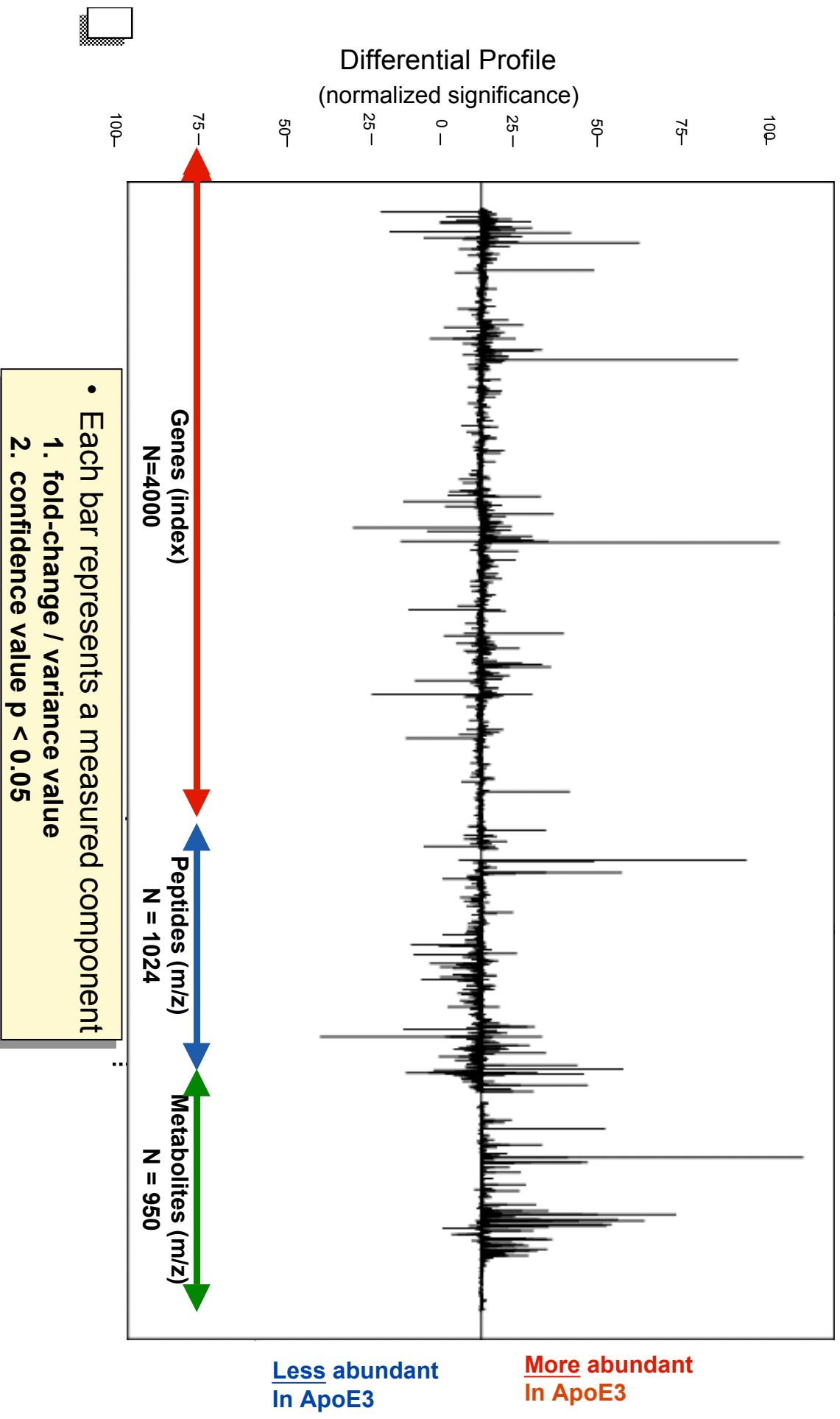


# Data Optimization: Quality Control

Data always includes:

- Error1      *Measurement*
- Error2      *Integrity*
- Noise      *Thermal*
- Randomness      *Markov Process*
- Variance      **Inherent Structure**

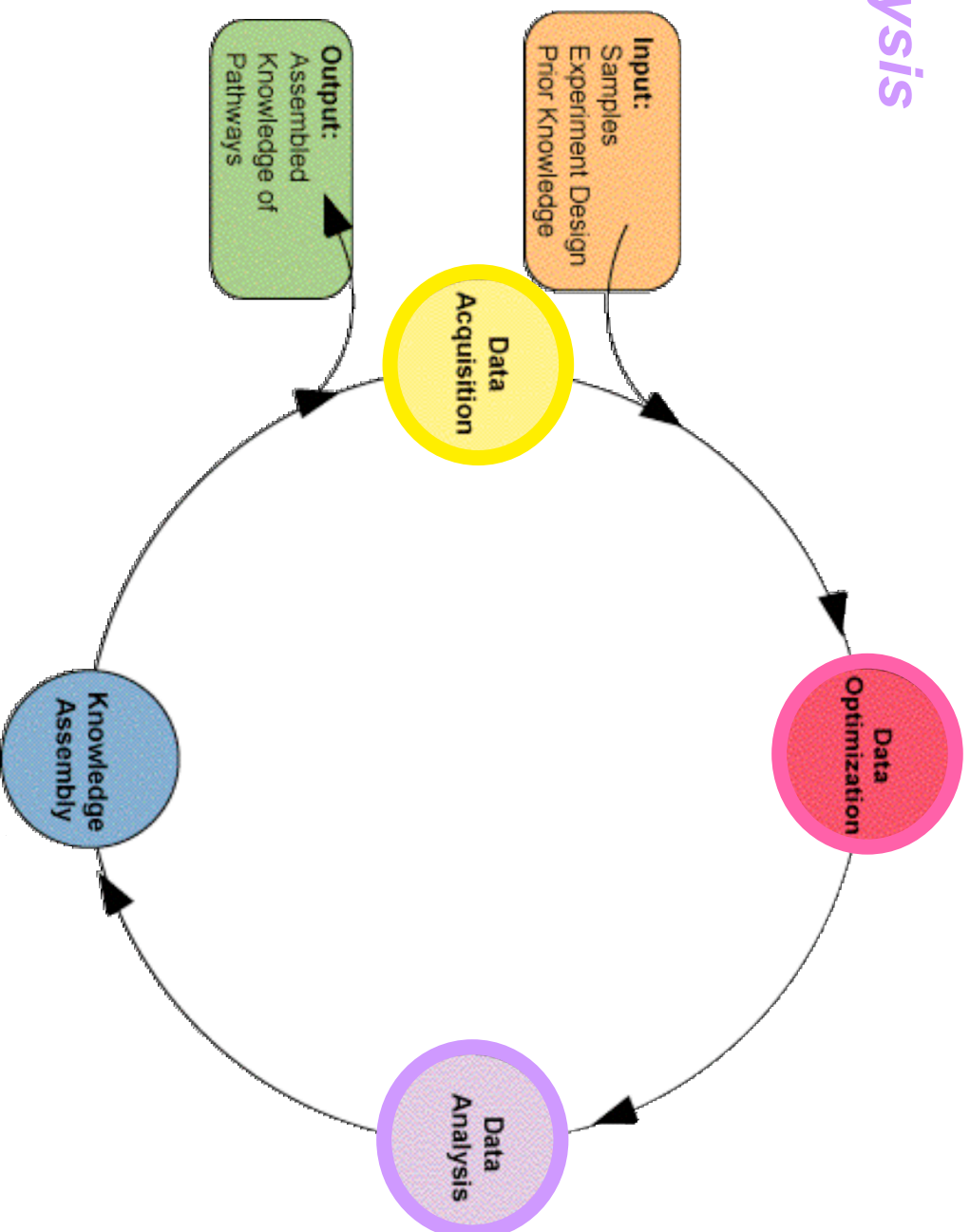
# Proteomic, Metabolomic, Genomic— Integrated Differential Expression Profile



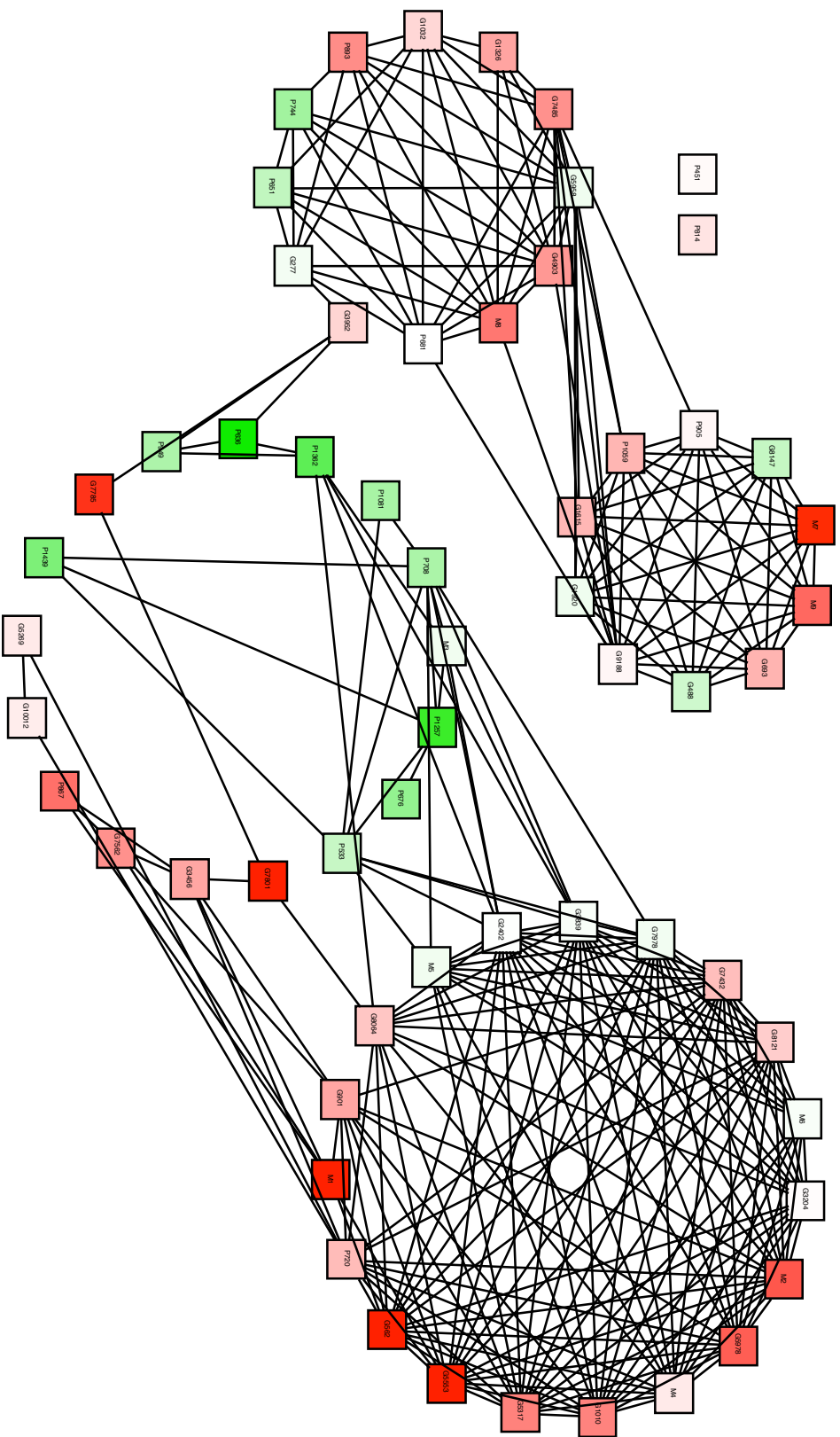


# BioSystematics™ at Beyond Genomics:

## Data Analysis

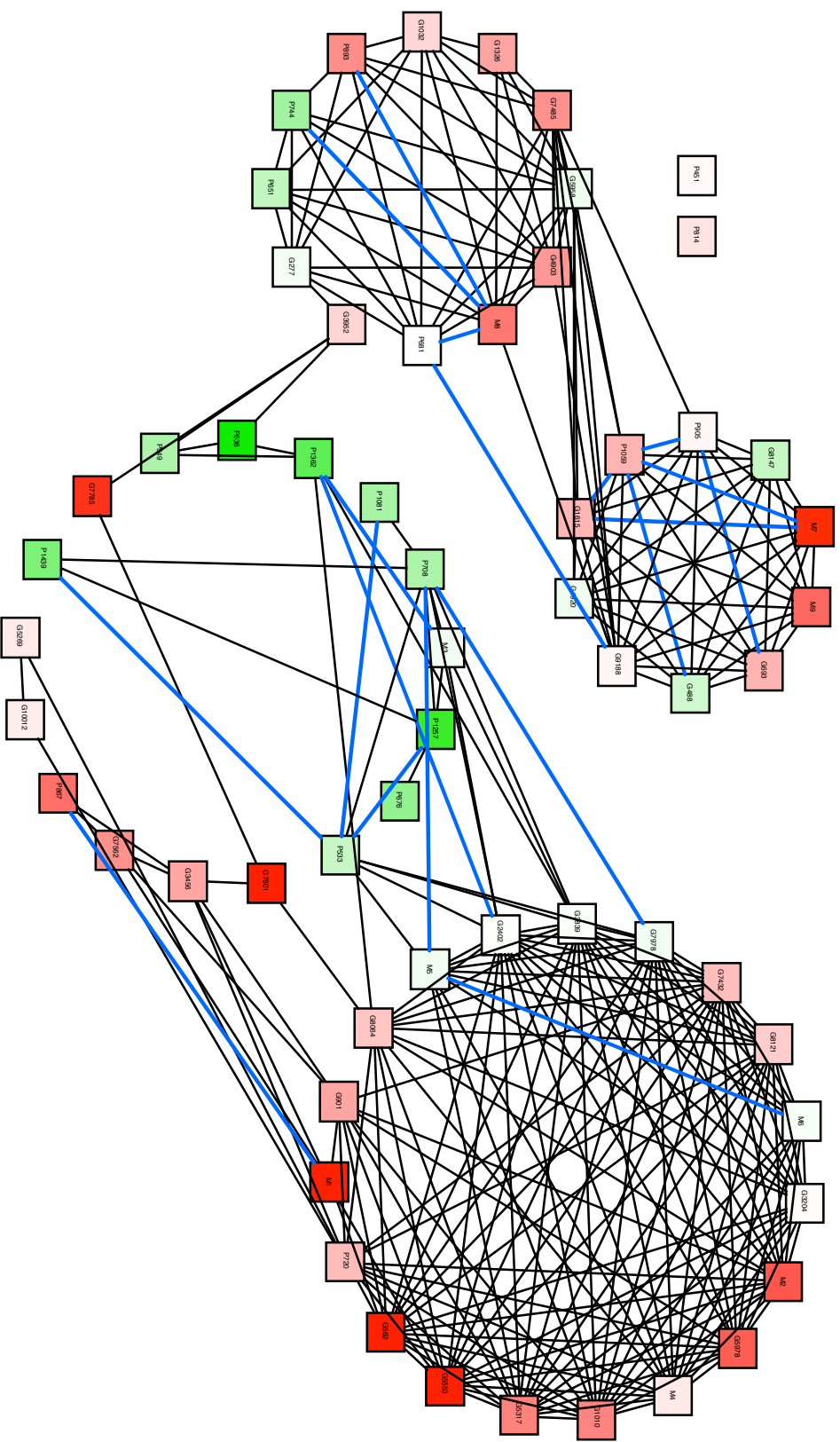


# Association Network Analysis



**Nodes connected based on high association using linear correlations, kernel PCA, or mutual information.**

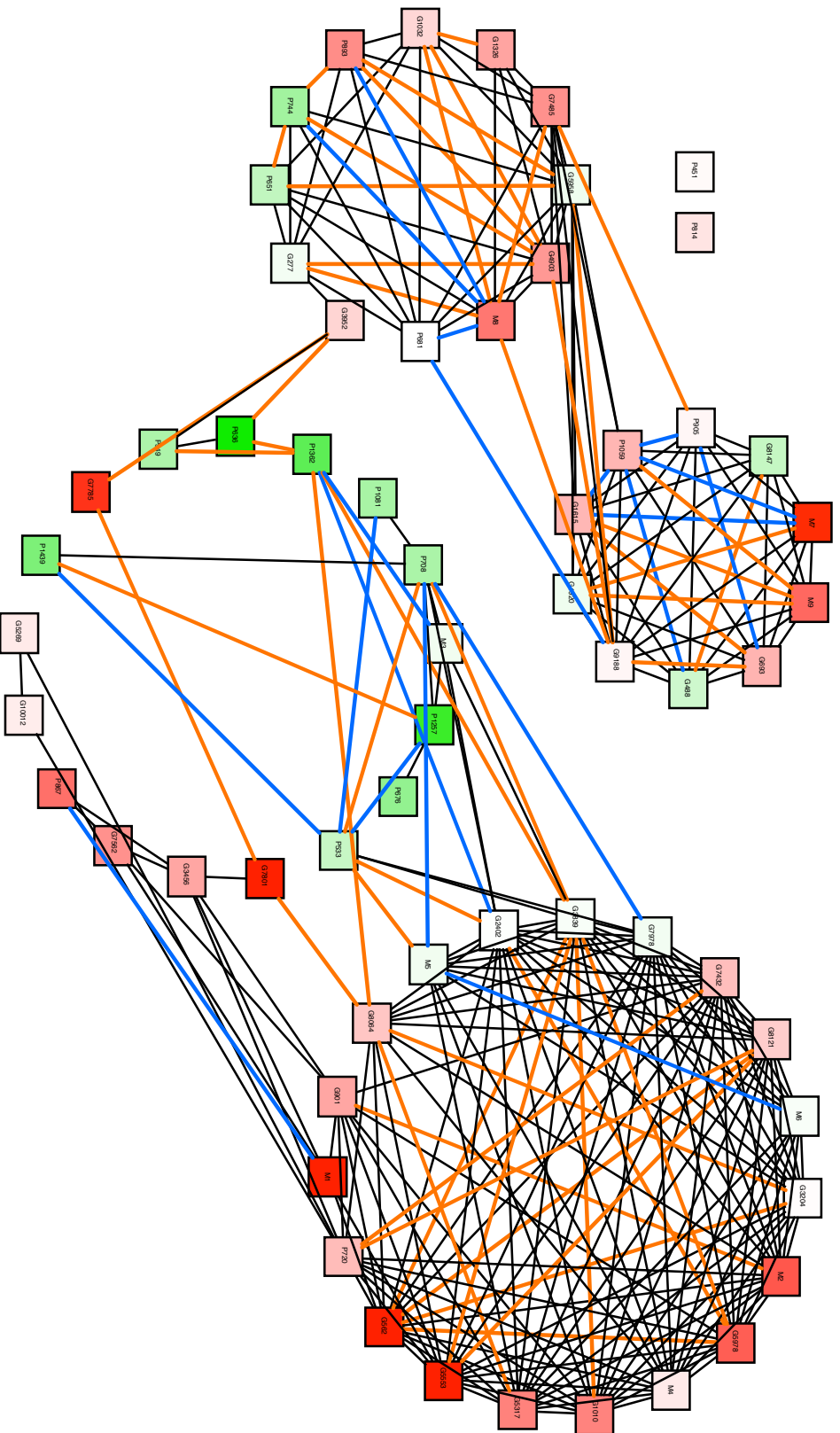
# Association Network Analysis



Annotation of edges using information from databases



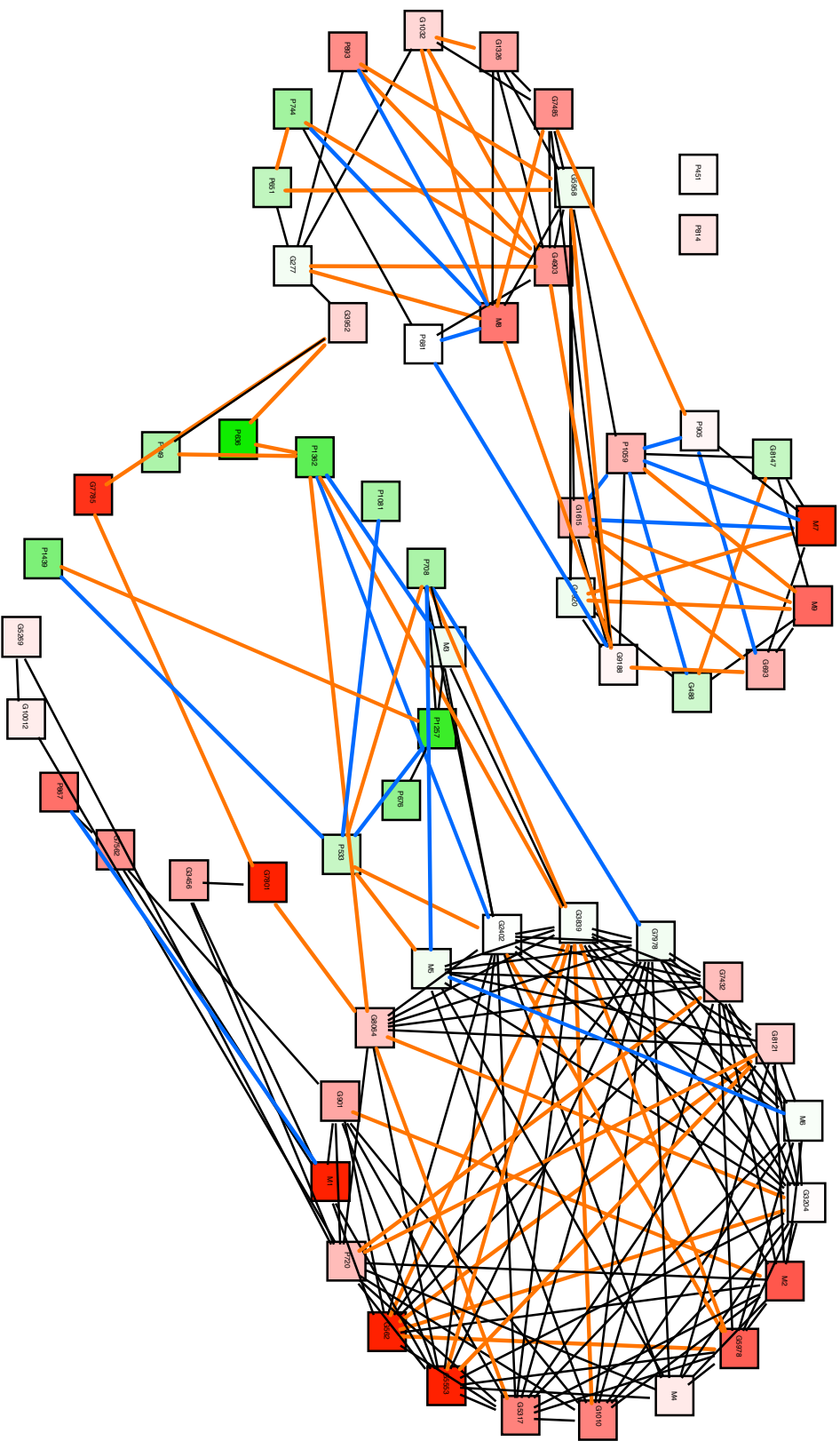
# Association Network Analysis



**Annotation of edges using information from literature:  
Co-occurrences of gene/protein/metabolite names in literature**

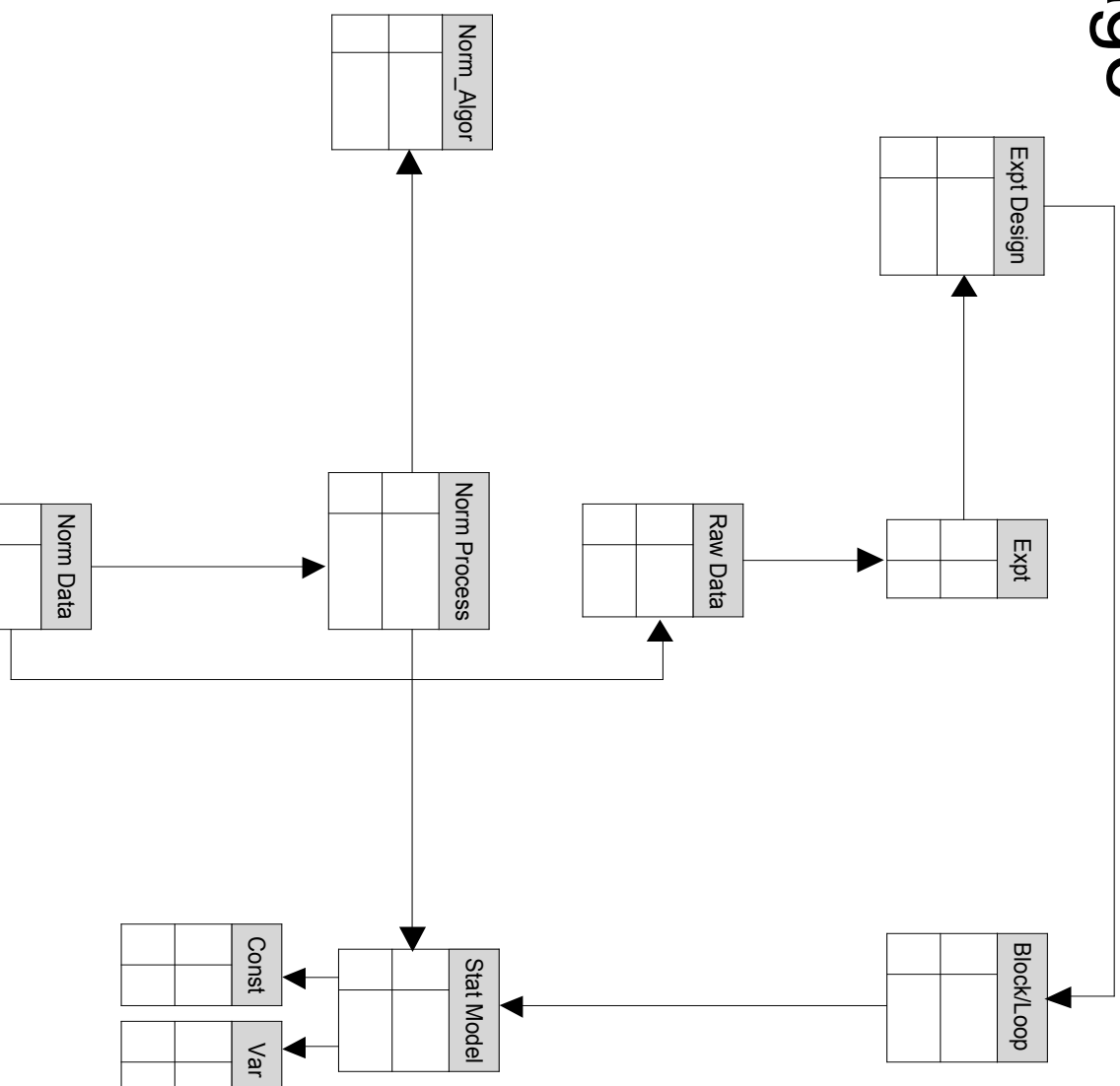


# Association Network Analysis



**We are left with known and unknown associations. If no further “cleanup” is possible, need to start model building and discovery from here (Network analysis, new hypothesis building, further experiments)**

# Statistical Model and Markup Language



## Pathway and Connection Visualization

- Bioinformatics + Pathways
  - Incellico
  - Genstruct
- Riken
- LSID – Common unique indexing system for mixing and merging heterogeneous data in both views and knowledge ... ala DASI
  - Ex: KEGG with LSID: Cmp(CAS), Rxn (EC), Enz (GB)

# Riken

about GSCOPE - Microsoft Internet Explorer

Address: <http://gscope.gsc.riken.go.jp/aboutGSCOPE.htm>

File Edit View Favorites Tools Help

Search Search Web Search Site PageRank Page Info Up Highlight

gsc RIKEN

**Introduction**  
Welcome to GSCOPE Knowledge Editor

**Sample data**  
> Pathway Mouse  
> Pathway Yeast  
> Ortholog  
> Yeast 2 hybrid

**Downloads**  
GSCOPE Knowledge Editor

**Online Manuals**  
GSCOPE Knowledge Editor

**About us**  
Staff Location

**Join us**

**Help Q&A**  
GSCOPE License Knowledge Editor FAQ

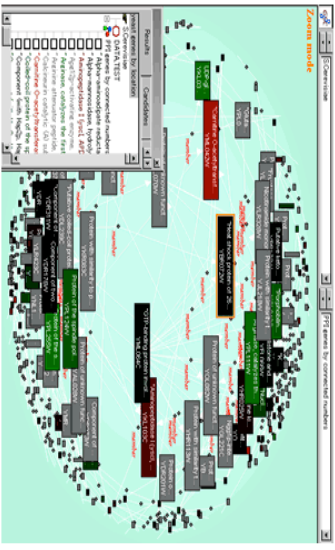
**Genomic Knowledge-base Software Gallery**

Home Introduction Sample data Downloads About us Help Q&A Site Map

**About GSCOPE**

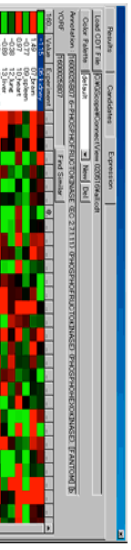
**Integrated viewer for biomolecular network graphs**

We present a new visualization software for understanding multi-linked complex biomolecular network graphs by combining advantages of fisheye conversion function and clipping of connected structures. In analysis of biomolecular network graphs, it helps us to understand microscopic relationships between molecules, as well as to grasp an overview of a large whole graph. Biomolecular networks extracted from various sort of biological data tend to be highly complicated and incomprehensible because it is often the case that found connections among the molecules are far from simple, but intercrossed with many other connection lines.



**Viewing and Analyzing expression data on biological pathways**

DNA microarrays are widely used to measure the expression levels of thousands of genes simultaneously. GSCOPE is designed for viewing and analyzing gene expression data in the context of biological pathways. GSCOPE has a function that filters thousands of gene expression data to extract statistically similar expression profiles. According to the researchers' requests in GSC, the above-mentioned filter is one of the most necessary function for integrated database of microarrays and useful for biological network finding work by expert biologists. GSCOPE will be also helpful for analysis of protein expression levels.



Internet

## **Data Analysis: CELL from Incellico**

- A principal technology for knowledge integration is Incellico's CELL, a database of relationships between databases
- CELL allows experimental findings to be constrained and enhanced by prior knowledge
- CELL supports data navigation and high-throughput data analysis.

## Databases Linked:

- OMIM
- SWISS-PROT-AC
- FLYBASE-ID
- PDB
- SWISS-PROT-ID
- GDB-ID
- PFAM-DESC
- SYMBOL
- GENE-Name
- PFAM-ID
- TAX-SCIENTIFIC-NAME
- GENPEPT
- PIR-ID, TAXID
- GENPEPT-ID
- PRINTS-DESC
- UNIGENE-ID
- GO
- PRINTS-ID
- CHROMOSOME
- GO-COMPONENT-DESC
- PRODOM-DESC
- CHROMOSOME-ARM
- GO-FUNCTION-DESC
- PRODOM-ID
- CHROMOSOME-BAND
- INTERPRO-DESC
- PROSITE-DESC
- CHROMOSOME-SUB-BAND
- INTERPRO-ID,
- PROSITE-ID
- EMBL/GENBANK
- KEYWORD,
- PUBMED
- EMBL/GENBANK-GI
- LOCUSLINK
- RATMAP-ID
- EMBL/GENBANK-LOCUS
- MEDLINE,
- REFSEQ
- EMBL/GENBANK
- VERSION
- MGI-ID
- SGD-ID
- EXPRESS-IN



# Incellico links GO Data to...

CELL Suite: Cross-Reference Clustering Results

```

root
├── LIGAND BINDING OR CARRIER
│   └── LIGAND BINDING OR CARRIER
│       └── PROTEIN SWISS-PROT-AC:P12710[...]
├── KINASE
│   ├── PROTEIN TYROSINE KINASE
│   │   ├── PROTEIN TYROSINE KINASE
│   │   │   └── PROTEIN SWISS-PROT-AC:008545[...]
│   │   └── PROTEIN TYROSINE KINASE
│   │       └── PROTEIN TYROSINE KINASE
│   │           └── PROTEIN SWISS-PROT-AC:008545[...]
│   └── TRANSFERASE, TRANSFERRING PHOSPHORUS-CONTAINING GR
│       └── PHOSPHOTRANSFERASE, ALCOHOL GROUP AS ACCEPTOR
│           └── PROTEIN TYROSINE KINASE
│               └── PROTEIN TYROSINE KINASE
│                   └── PROTEIN SWISS-PROT-AC:008545[...]
├── INOSITOLPHOSPHATIDYLINOSITOL PHOSPHATASE
│   ├── INOSITOLPHOSPHATIDYLINOSITOL PHOSPHATASE
│   │   └── PROTEIN SWISS-PROT-AC:008545[...]
│   └── PROTEIN SWISS-PROT-AC:056023[...]
├── HYDROLASE, ACTING ON ESTER BONDS
│   ├── PHOSPHORIC MONOESTER HYDROLASE
│   │   ├── INOSITOLPHOSPHATIDYLINOSITOL PHOSPHATASE
│   │   │   └── PROTEIN SWISS-PROT-AC:056023[...]
│   │   └── INOSITOLPHOSPHATIDYLINOSITOL PHOSPHATASE
│   │       └── PROTEIN SWISS-PROT-AC:056023[...]
│   └── LIGASE, FORMING CARBON-NITROGEN BONDS
│       ├── ACID-D-AMINO ACID LIGASE
│       │   └── UBIQUITIN--PROTEIN LIGASE
│       │       └── PROTEIN SWISS-PROT-AC:008759[...]
│       └── UBIQUITIN--PROTEIN LIGASE
│           └── PROTEIN SWISS-PROT-AC:070252[...]
├── OXIDOREDUCTASE, ACTING ON PAIRED DONORS, WITH INCORP
│   ├── OXIDOREDUCTASE, ACTING ON PAIRED DONORS, WITH INCO
│   │   ├── HEME OXYGENASE (DECYCLIZING)
│   │   │   └── HEME OXYGENASE (DECYCLIZING)
│   │   │       └── PROTEIN SWISS-PROT-AC:070252[...]
│   │   └── HEME OXYGENASE (DECYCLIZING)
│   │       └── PROTEIN SWISS-PROT-AC:070252[...]
│   └── OXIDOREDUCTASE, ACTING ON PAIRED DONORS, WITH INCO
│       ├── HEME OXYGENASE (DECYCLIZING)
│       │   ├── HEME OXYGENASE (DECYCLIZING)
│       │   │   └── PROTEIN SWISS-PROT-AC:070252[...]
│       │   └── HEME OXYGENASE (DECYCLIZING)
│       │       └── PROTEIN SWISS-PROT-AC:070252[...]
│       └── HEME OXYGENASE (DECYCLIZING)
│           └── PROTEIN SWISS-PROT-AC:070252[...]

```

```

root
├── PROTEIN SWISS-PROT-AC:P12710[...]
├── 2.7.-.-
│   └── PHOSPHOTRANSFERASE, ALCOHOL GROUP AS ACCEPTOR
│       ├── PROTEIN KINASE
│       │   ├── PROTEIN TYROSINE KINASE
│       │   │   └── PROTEIN SWISS-PROT-AC:008545[...]
│       │   └── PROTEIN TYROSINE KINASE
│       │       └── PROTEIN SWISS-PROT-AC:P12710[...]
│       └── PROTEIN SWISS-PROT-AC:P12710[...]
├── 3.1.-.-
│   ├── 3.1.3.-
│   │   ├── 3.1.3.25
│   │   │   └── 3.1.3.25
│   │   └── PROTEIN SWISS-PROT-AC:P12710[...]
│   └── PROTEIN SWISS-PROT-AC:056023[...]
├── 6.3.-.-
│   ├── 6.3.2.-
│   │   └── 6.3.2.19
│   │       └── PROTEIN SWISS-PROT-AC:008759[...]
│   └── PROTEIN SWISS-PROT-AC:P12710[...]
├── 1.14.-.-
│   ├── 1.14.99.-
│   │   ├── 1.14.99.3
│   │   │   └── 1.14.99.3
│   │   └── PROTEIN SWISS-PROT-AC:070252[...]
│   └── 1.14.99.3
│       ├── 1.14.99.-
│       │   ├── 1.14.99.3
│       │   │   └── 1.14.99.3
│       │   └── PROTEIN SWISS-PROT-AC:070252[...]
│       └── 1.14.99.3
│           └── PROTEIN SWISS-PROT-AC:070252[...]

```

Save Left Tree...

Save to JPEG...

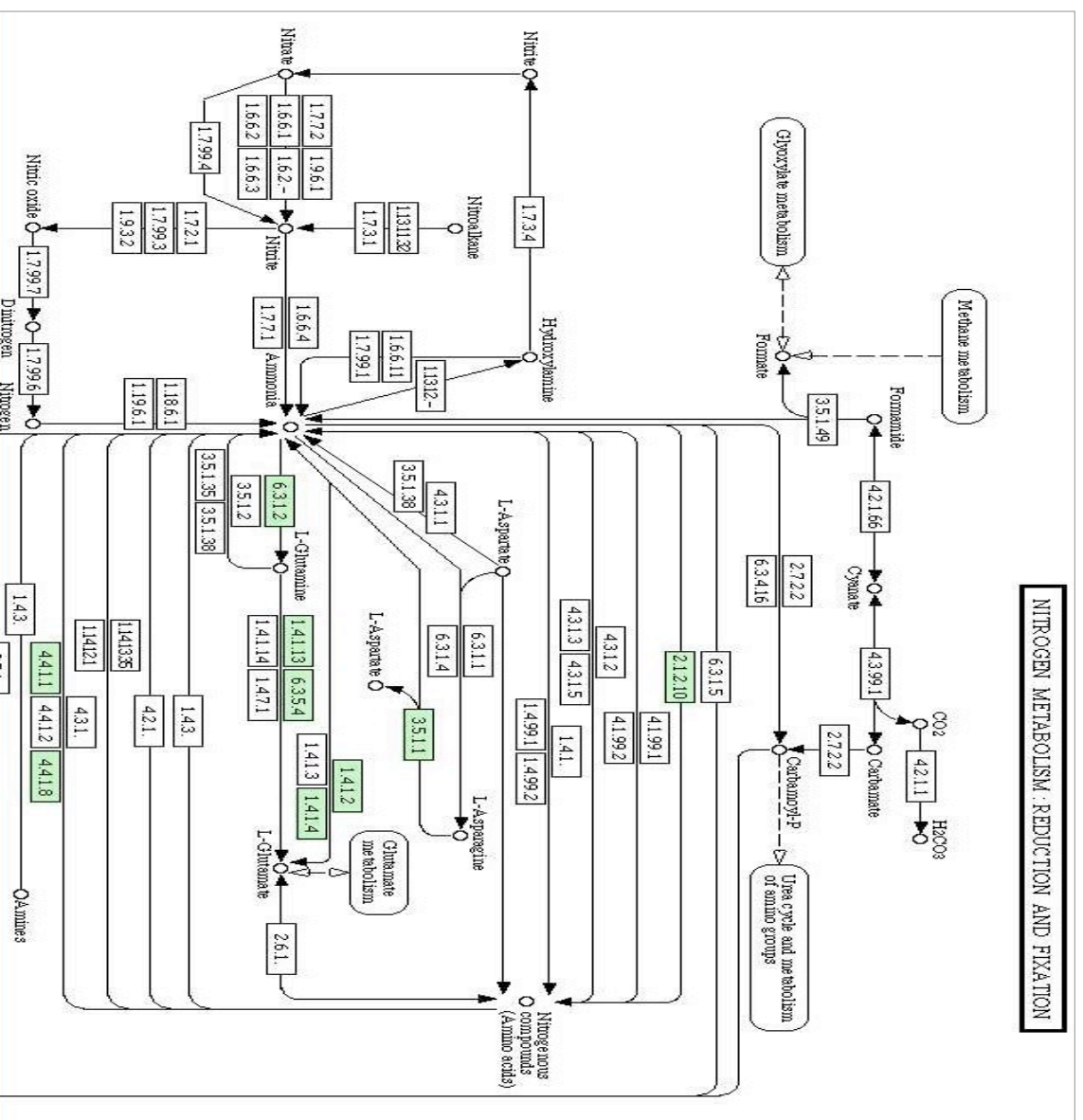
Save to HTML...

Close

Help

Save Right Tree...

# KEGG Metabolic Pathways



# Cell Search: Go Molecular Function: Death Receptor

**CELL suite Entity Browser**

Previously Viewed Entities: CLASSIFICATION GO-FUNCTION-DESC:DEATH [...] (Viewed 2 times) | Previous Entity | Next Entity

CLASSIFICATION GO-FUNCTION-DESC:TRANSMEMEM[...] (GO:0004888)

CLASSIFICATION GO-FUNCTION-DESC:DEATH REL[...] (GO:0005035)

CLASSIFICATION GO-FUNCTION-DESC:DEATH REL[...] (GO:0005040)

CLASSIFICATION GO-FUNCTION-DESC:DEATH REL[...] (GO:0005037)

CLASSIFICATION GO-FUNCTION-DESC:DEATH REL[...] (GO:0005038)

CLASSIFICATION GO-FUNCTION-DESC:DEATH REL[...] (GO:0005039)

CLASSIFICATION GO-FUNCTION-DESC:TUMOR NE[...] (GO:0005031)

CLASSIFICATION GO-FUNCTION-DESC:DEATH REL[...] (GO:0005039)

*Entity Browser: CELL Suite by InoeHiro, Inc. (www.inoeHiro.com)*

<no attributes>

24

Help

# Drill for Tumor Necrosis Factor, a specific Death Receptor

CELL Suite Entity Browser

Previously Viewed Entities

CLASSIFICATION GO-FUNCTION-DESC:TUMOR [...] Previous Entity Next Entity

CLASSIFICATION GO-FUNCTION-DESC:DEATH REL[...] GO:0005035

CLASSIFICATION GO-FUNCTION-DESC:TUMOR NE[...] GO:0005027

CLASSIFICATION GO-FUNCTION-DESC:TUMOR NE[...] GO:0005031

LOCUS LOCUSLINK-DESC:TNFRSF1B: LOCUSLINK:7133

PROTEIN SWISSPROT-DESC:TNF receptor SWISS-PROT-AC:P39428 SWISS-PROT-ID:TRAF1\_MOUSE [More.]

LOCUS LOCUSLINK-DESC:TNFRSF1A: LOCUSLINK:7132

PROTEIN SWISSPROT-DESC:DG17 prote DICTYDB-ID:DD02010 PIR-ID:A29361 [More.]

LOCUS LOCUSLINK-DESC:TNFRSF1A: LOCUSLINK:7132

PROTEIN SWISSPROT-DESC:TNF receptor SWISS-PROT-AC:Q13077 SWISS-PROT-ID:TRAF1\_HUMAN [More.]

Entity Browser: CELL Suite by Inetllo, Inc. (www.inetllo.com)

<no attributes>


15


Help


# Choose the Human Protein and view the full range of associated data


**CELL Suite Entity Browser**


Previously Viewed Entities: [PROTEIN SWISS-PROT-AC:Q13077L.] [Viewed 3 times] | Previous Entity | Next Entity


**CLASSIFICATION**  
 GO-FUNCTION-DESC:TUMOR NE[...]  
 GO:0005031  
[\[More..\]](#)


**SYMB**  
 **SYMBOL**  
 SYMBOL:&BFR;FTZ-F1  
 SYMBOL:&BGR;4-TUBULIN  
 SYMBOL:0123/09  
[\[More..\]](#)


**TAXONOMY**  
 TAX-OTHER-NAMES:human, ma  
 TAX-NAME:HOMO SAPIENS  
 TAXID:9606  
[\[More..\]](#)


**EXPRESS-IN**  
 EXPRESS-IN:LYMPHOMA


**LITERATURE**  
 MEDLINE:95163092  
 PUBMED:7859281  
[\[More..\]](#)


**PROTEIN**  
 SWISSPROT-DESC:TNF recept  
 SWISS-PROT-AC:Q13077  
 SWISS-PROT-ID:TRAF1\_HUMAN  
[\[More..\]](#)

**GENE-PRODUCT**  
 UNIGENE-DESC:TNF receptor  
 UNIGENE-ID:H8.2134

**GENE-PRODUCT**  
 GDB-ID:6268627


**PROTEIN**  
 GENPEPT-GI:5032193  
 REFSEQ-PROT:NP\_005649


**LOCUS**  
 LOCUSLINK-DESC:TRAF1: TNF  
 LOCUSLINK:7185

**GENE**  
 OMIM-DESC:TNF receptor-as  
 OMIM:601711

Entity Browser: CELL Suite by Inosilico, Inc. ([www.inosilico.com](http://www.inosilico.com))

Attributes: SWISSPROT-DESC:TNF receptor associated factor 1 (TRAF1) (Epstein-Barr virus-induced protein 6);

7 

7 


Help


# Horizontal Scrolling Allows Searching the Breadth of Data


**CELL Suite Entity Browser**


Previously Viewed Entities **PROTEIN SWISS-PROT-AC:Q13077[...]** [Viewed 4 times] Previous Entity Next Entity


5
2


**MOTIF**  

**INTERPRO-DESC: MATH**  
 INTERPRO-ID:IPR002083


**MOTIF**  

**INTERPRO-DESC: TRAF**  
 INTERPRO-ID:IPR003007


**MOTIF**  

**PFAM-DESC: MATH**  
 PFAM-ID:PF00917


**MOTIF**  

**SMART-ID: SM000061**  
 SMART-NAME: MATH


**CLASSIFICATION**  

 KEYWORD: COILED COIL


**PROTEIN**  

 SWISSPROT-DESC: TNF recept  
 SWISS-PROT-AC: Q13077  
 SWISS-PROT-ID: TRAF1\_HUMAN  
 [More...]

**GENE-PRODUCT**  

 UNIGENE-DESC: Tnfr receptor  
 MGI-ID: 101836  
 UNIGENE-ID: MIM: 12898  
 [More...]

**NUC-SEQUENCE**  

 EMBLGENBANK-DESC: Homo sa  
 EMBLGENBANK-GI: 675461  
 EMBLGENBANK-LOCUS: HSU192L[...]  
 [More...]

**GENE-PRODUCT**  

 UNIGENE-DESC: ESTs, Highly  
 UNIGENE-ID: BT: 9083

**PROTEIN**  

 GENPEPT-GI: 675462  
 GENPEPT-AAA62309

**GENE-PRODUCT**  

 UNIGENE-DESC: ESTs, Highly  
 UNIGENE-ID: MIM: 123297

Entity Browser: CELL Suite by Koezllioo, Inc. (www.koezllioo.com)

Attributes: SWISSPROT-DESC: TNF receptor associated factor 1 (TRAF1) (Epstein-Barr virus-induced protein 8) ;

Help

# Locus Link accessible via Nucleic Sequence

CELL suite Entity Browser


Previously Viewed Entities

GENE OMIM:601711[...]


Previous Entity

Next Entity




 **LOCUS**  
LOCUSLINK-DESC:TRAF1: TNF  
LOCUSLINK:7185


 **PHYSICAL-LOC**  
CHROMOSOME-BAND:9q-q1[...]  
CHROMOSOME-BAND:9q-q1[...]

 **PROTEIN**  
SWISSPROT-DESC:TNF\_recept  
SWISS-PROT-AC:Q13077  
SWISS-PROT-ID:TRAF1\_HUMAN  
[More...]

 **PHYSICAL-LOC**  
CHROMOSOME-BAND:9q-q1[...]

 **PHYSICAL-LOC**  
CHROMOSOME:9606-9

 **GENE**  
OMIM-DESC:TNF\_receptor-as  
OMIM:601711

 **NUC-SEQUENCE**  
EMBLGENBANK-DESC:Homo sa  
EMBLGENBANK-GI:5032192  
EMBLGENBANK-LOCUS:NM\_005[...]  
[More...]

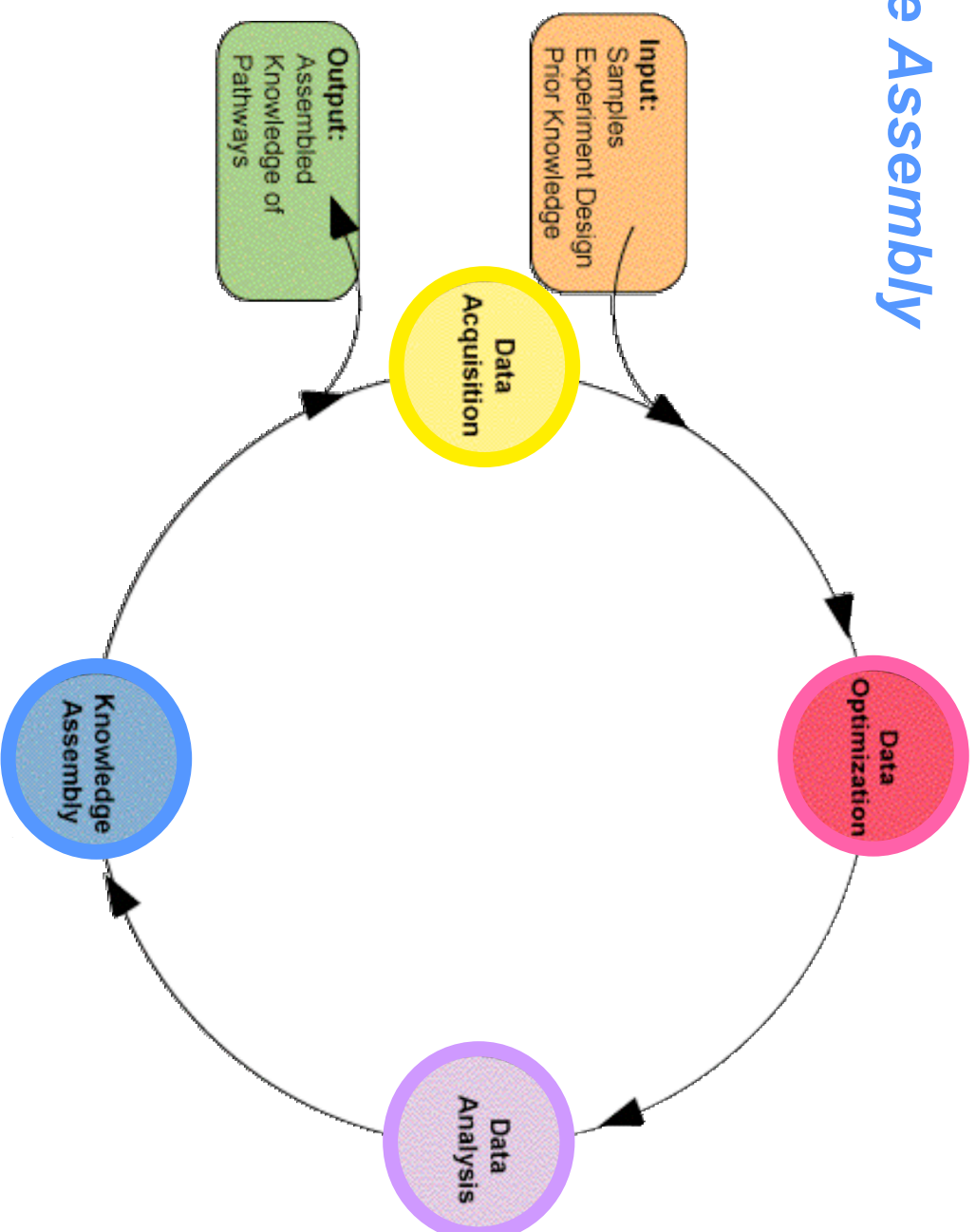
Entity Browser: CELL Suite by InCellBio, Inc. (www.incellbio.com)

Attributes: OMIM-DESC:TNF\_receptor-associated factor 2;

Help

# BioSystematics™ at Beyond Genomics:

## Knowledge Assembly






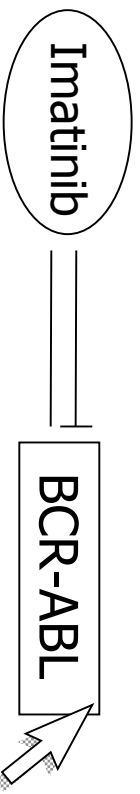





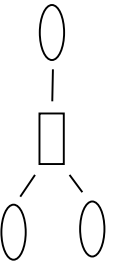
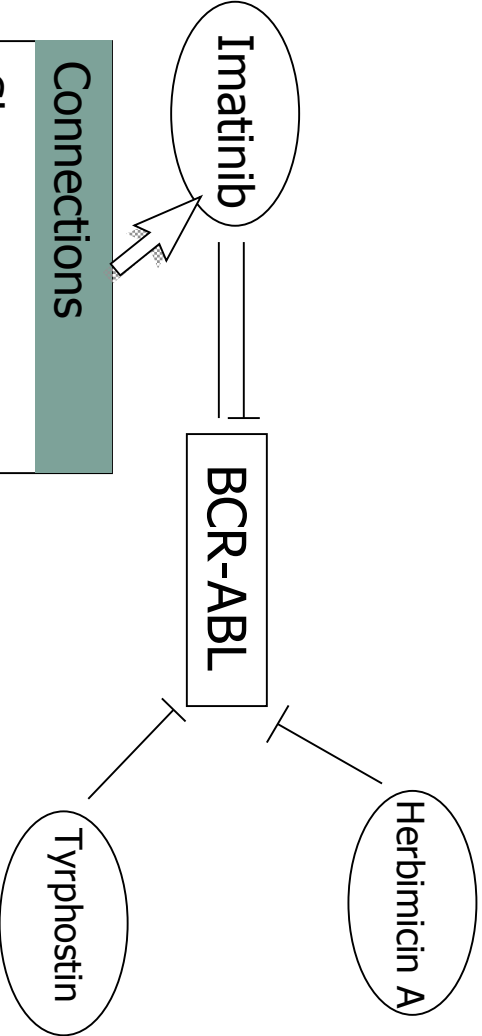
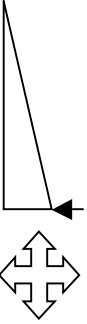
# Knowledge Representation / Assembly: (Genstruct)

- Provides Expressivity to Biology; many complex semantics that WILL change in time, that need to be captured!
- Provides a software environment for revealing the logic of biology
- Uses technology that renders connections of any type (logic bytes) in an environment in which scientists can:
  - Navigate connections between biological objects and concepts
  - Query the connections
  - Reason computationally

# Discovery Environment

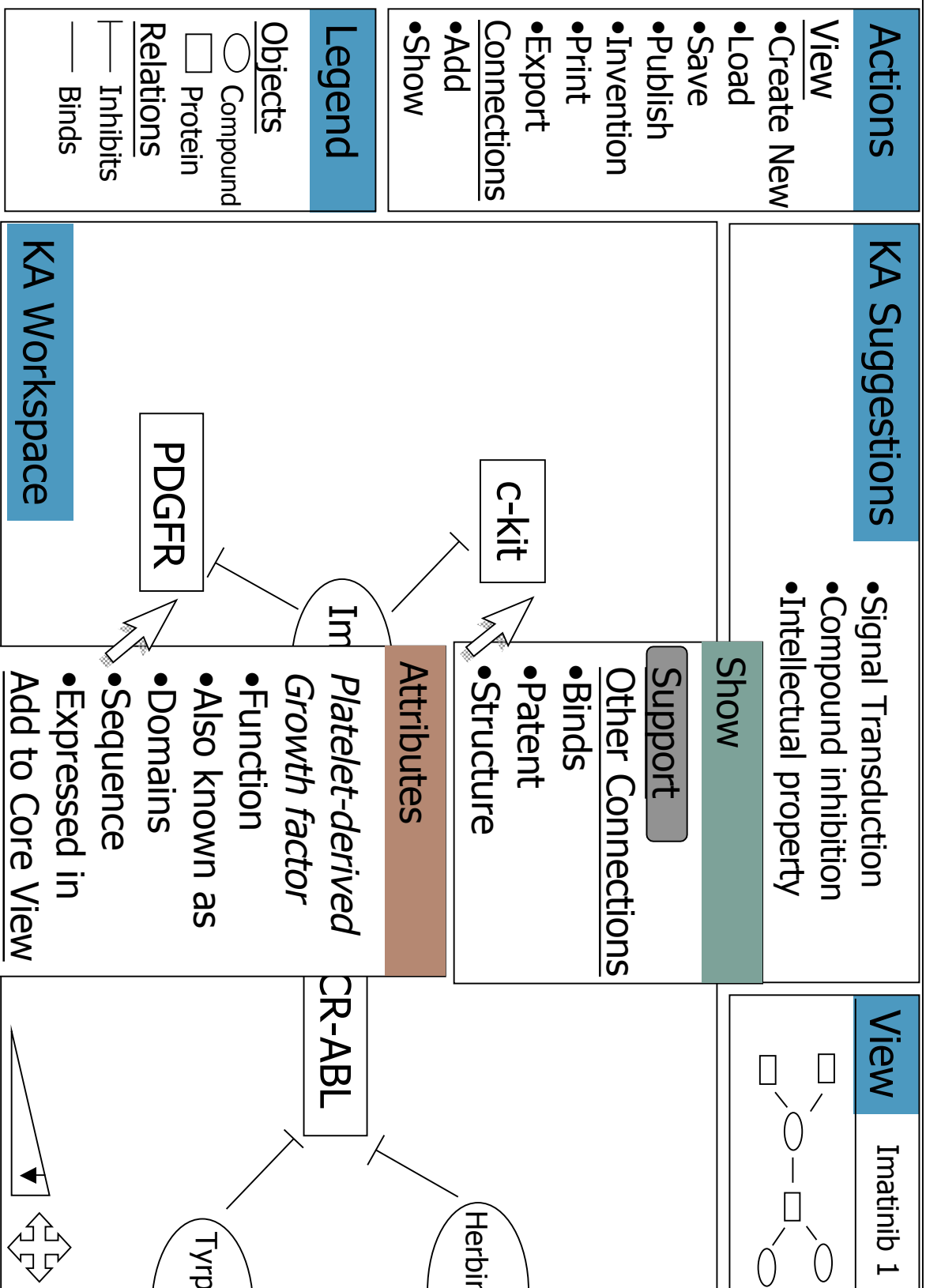
<b>Actions</b> <u>View</u> <ul style="list-style-type: none"> <li>• Create New</li> <li>• Load</li> <li>• Save</li> <li>• Publish</li> <li>• Invention</li> <li>• Print</li> <li>• Export</li> </ul> <u>Connections</u> <ul style="list-style-type: none"> <li>• Add</li> <li>• Show</li> </ul>	<b>KA Suggestions</b> <ul style="list-style-type: none"> <li>• Compound inhibition</li> <li>• HTS</li> <li>• Intellectual property</li> </ul>	<b>View</b> Imatinib 1 
<b>Legend</b> <u>Objects</u>  Compound  Protein <u>Relations</u>  Inhibits  Binds	<b>KA Workspace</b>  <div data-bbox="284 1249 706 1680" style="border: 1px solid black; padding: 5px;"> <b>Connections</b>  <u>Show</u>  <ul style="list-style-type: none"> <li>• Phosphorylates</li> <li>• Activates</li> <li>• <b>Inhibited by</b></li> </ul> <u>Create new</u> </div> 	

# Discovery Environment

<b>Actions</b> <ul style="list-style-type: none"> <li>• Create New</li> <li>• Load</li> <li>• Save</li> <li>• Publish</li> <li>• Invention</li> <li>• Print</li> <li>• Export</li> </ul>	<b>KA Suggestions</b> <ul style="list-style-type: none"> <li>• Compound inhibition</li> <li>• HTS</li> <li>• Intellectual property</li> </ul>	<b>View</b> Imatinib 1 
<b>View</b> <ul style="list-style-type: none"> <li>• Create New</li> <li>• Load</li> <li>• Save</li> <li>• Publish</li> <li>• Invention</li> <li>• Print</li> <li>• Export</li> </ul> <b>Connections</b> <ul style="list-style-type: none"> <li>• Add</li> <li>• Show</li> </ul>	<div style="text-align: center;">  </div>	
<b>Legend</b> <b>Objects</b> <ul style="list-style-type: none"> <li>○ Compound</li> <li>□ Protein</li> </ul> <b>Relations</b> <ul style="list-style-type: none"> <li>— Inhibits</li> <li>— Binds</li> </ul>	<b>Connections</b> <ul style="list-style-type: none"> <li>• Show</li> <li>• HTS</li> <li>• <b>Inhibits</b></li> <li>• Binds</li> <li>• Create new</li> </ul>	

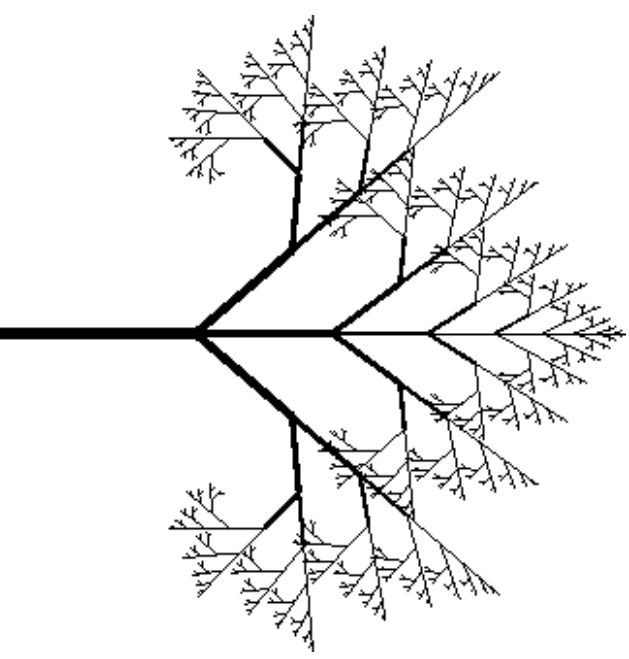


# Discovery Environment



# Computational Reasoning in the Knowledge Base

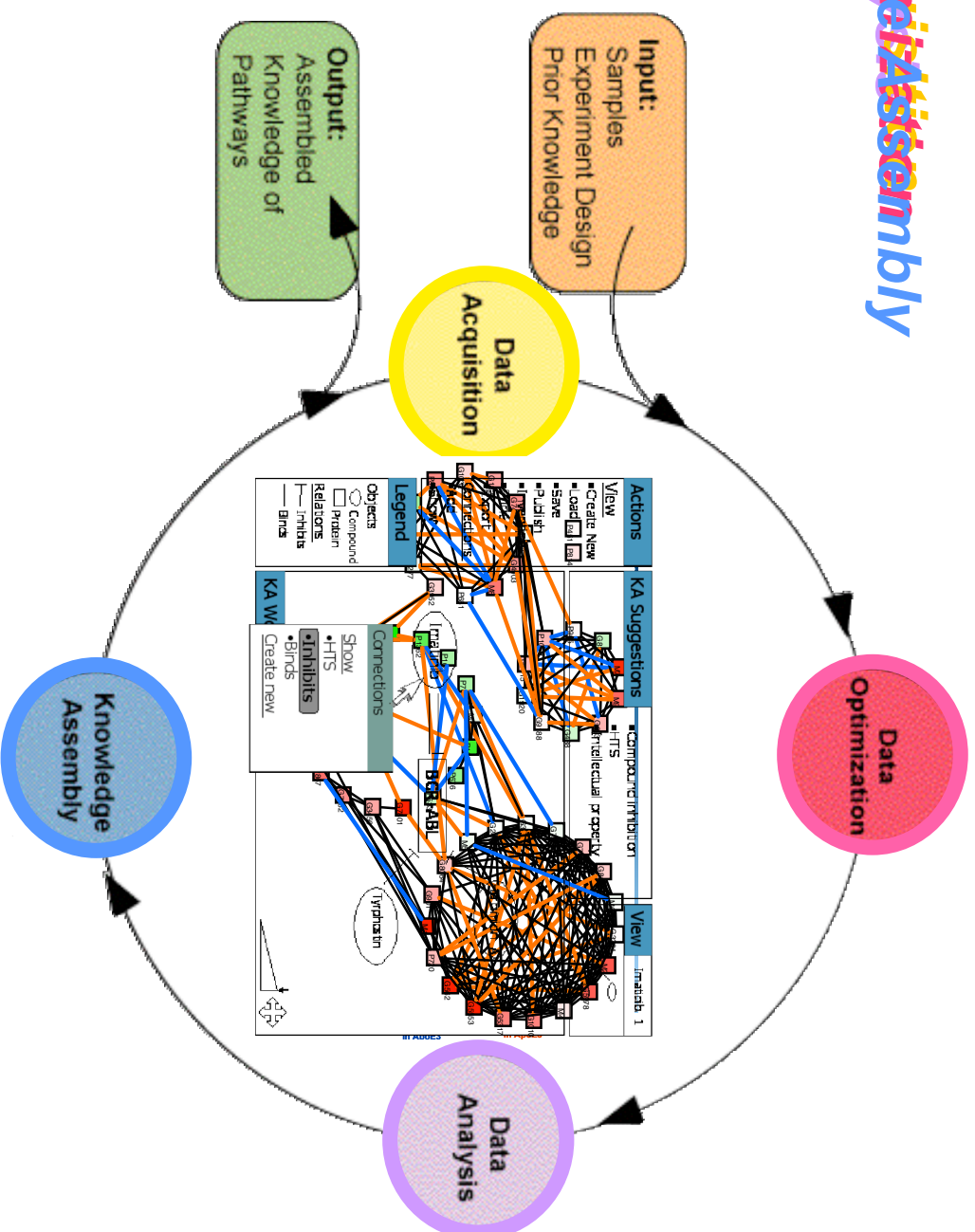
- Ontologies: biologic axioms, normalized nomenclature
- Algorithms: Relevance network analysis, Bayesian nets
- Knowledge mining: Graph equivalence, Arborecence, Hamiltonian paths and subgraph matching
- Intelligent Agents: User-defined parameters/algorithms search assembled knowledge



*arborecence*

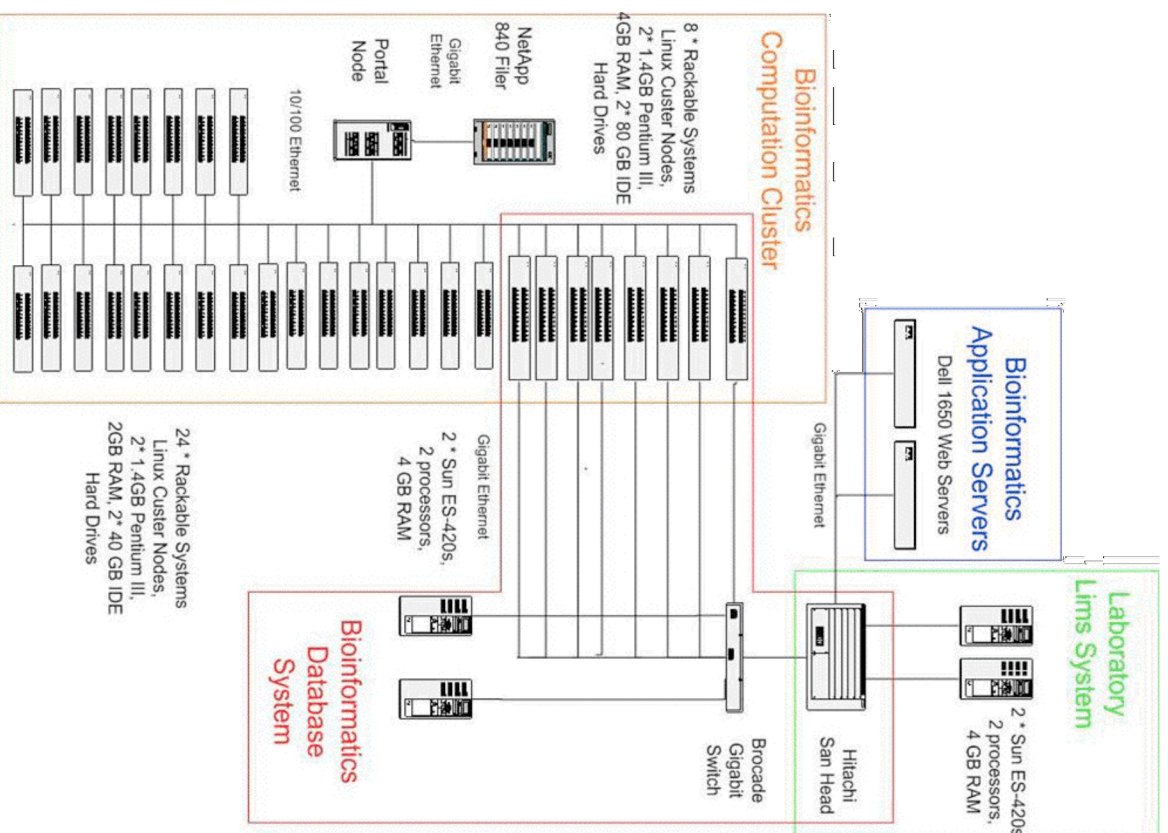
# The Integrated Flow:

*Knowledge Assembly*



# BioSystematics™ at Beyond Genomics:

## Hardware Platform



## Issues raised by Use of Clusters

- Clustered computing solutions are often a good way to cope with the need for increased computational power, but . . .
- What a world needs is a good cluster management tool
- The problem of “by processor licensing”



# Reiteration

- Moving Beyond the Genome Utopia: Systems Biology requires Bioinformatics to take its tools, expertise and agenda to a higher level
- Standards to Cope with Proliferation of Platforms and Vendors
- Lims Systems and Pipeline Management
- Data Analysis Standards
- Visualization Tools
- Working with Clusters, Avoiding Bankruptcy

# Acknowledgements

- Eric Neumann
- Matej Oresic
- Tom Plasterer
- Carl Ruel
- Stacey Horrigan
- Genstruct: Jeffrey Thomas
- Incellico: Tre-Tin Lee