

# Harmony from a users perspective

From biology to data and back again,  
the modeling lane

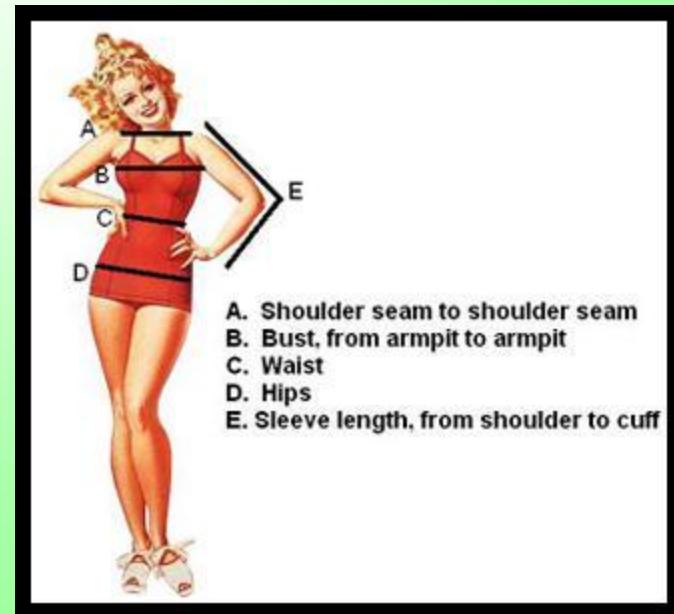
Chris Evelo  
Department of  
Bioinformatics - BiGCaT  
Maastricht University

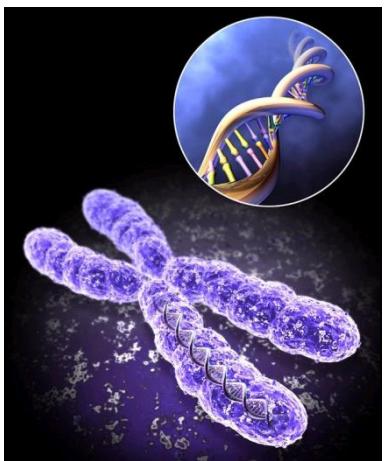
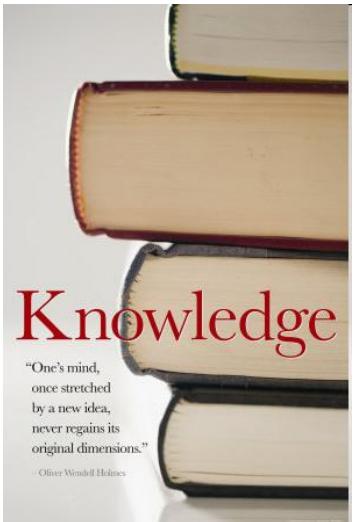


# Modeling

## Understanding what we measure

Using what we already know





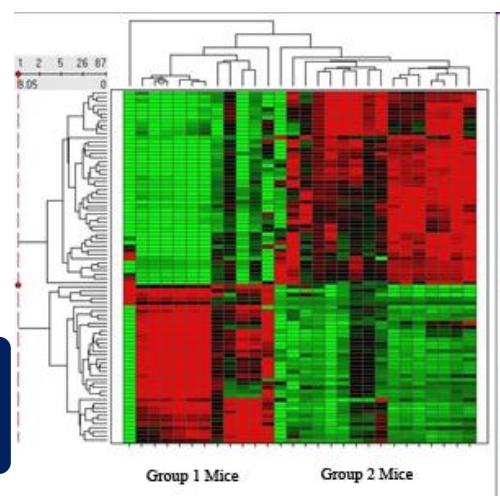
Existing  
knowledge



Genetic  
Results



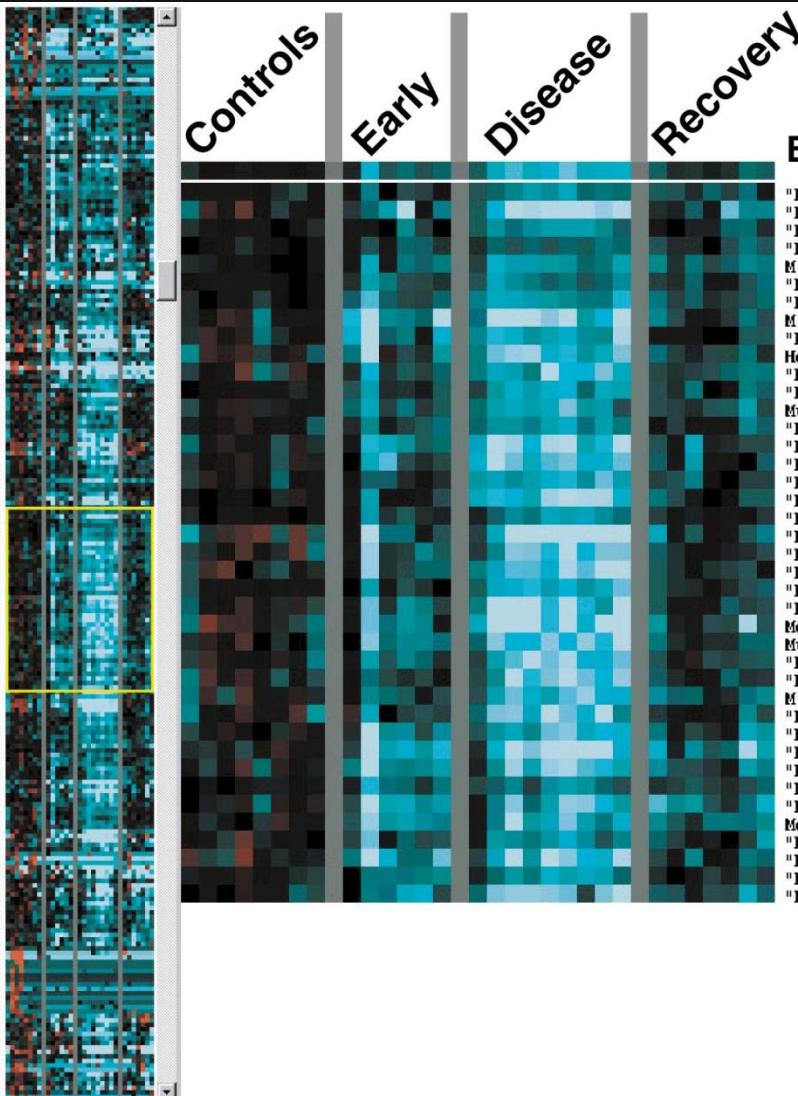
Genomic  
Results



There is a lot  
of knowledge  
to structure



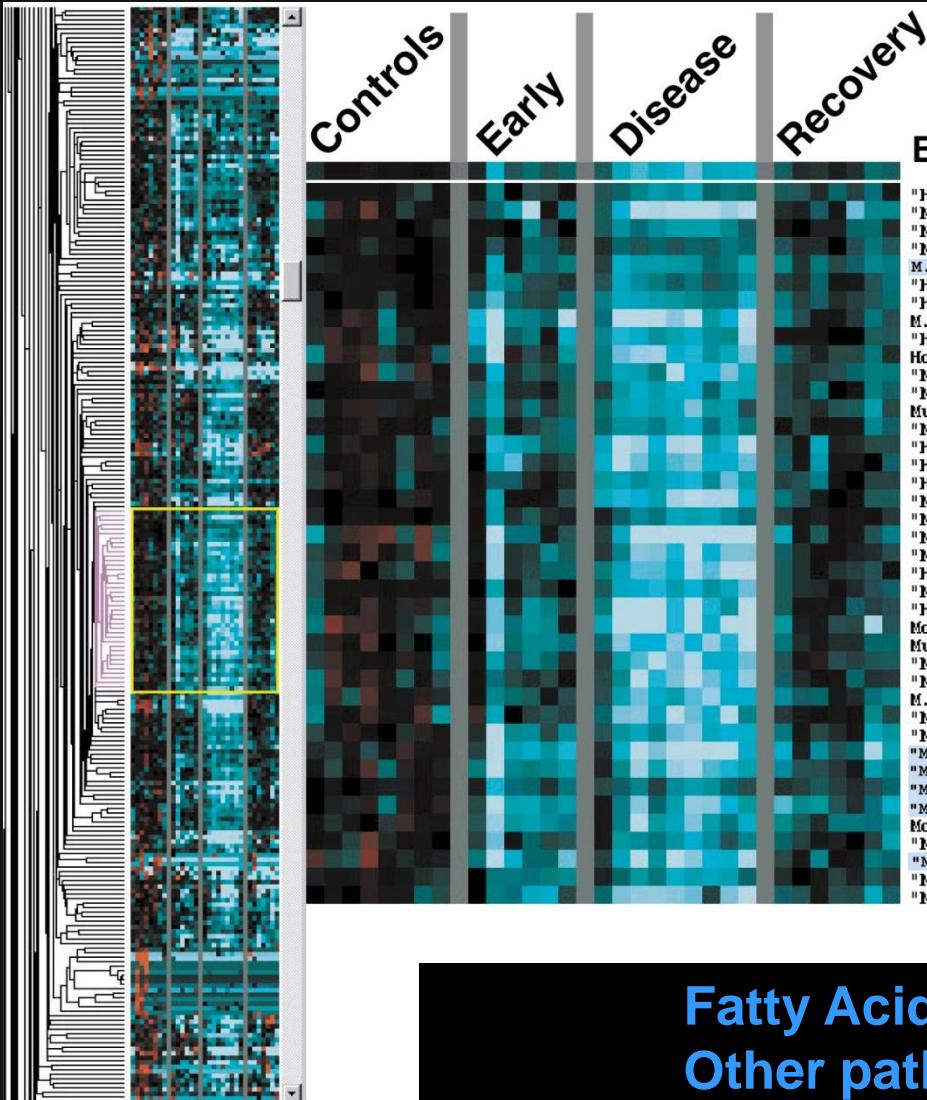
# Cardiomyopathy: Downregulated genes



## BLAST Definitions

"Homologous to sw 007021: PRE-MRNA SPLICING FACTOR SF2, P32 SUBUNIT PRECURSOR (GCIQ-R PROTEIN)  
"Mus musculus proteasome activator PA28 alpha subunit mRNA, complete cds"  
"Mus musculus cdc37 homolog mRNA, complete cds"  
"Mus musculus ornithine decarboxylase antizyme gene, complete cds"  
M.musculus mRNA for carnitine acetyltransferase  
"Homologous to sw 000779: CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE (EC 3.6.1.36)  
"Homologous to sw P11507: CALCIUM-TRANSPORTING ATPASE ENDOPLASMIC RETICULUM TYPE (EC 3.6.1.38)  
M.musculus ENO3 mRNA for enolase beta subunit  
"Homologous to sw P47858: 6-PHOSPHOFRUCTOKINASE, MUSCLE TYPE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE)  
"Homologous to sw P23327: SARCOPLASMIC RETICULUM HISTIDINE-RICH CALCIUM-BINDING PROTEIN PRECURS  
"Mouse AE3 mRNA, complete cds"  
"M.musculus glucose transporter 2 mRNA, complete cds"  
Mus musculus aspartate aminotransferase gene 5'-flank and exon 1  
"Mus musculus thioredoxin-dependent peroxide reductase (txn) mRNA, complete cds"  
"Homologous to sw P47858: 6-PHOSPHOFRUCTOKINASE, MUSCLE TYPE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE)  
"Homologous to sw P11508: CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE (EC 3.6.1.38)  
"Homologous to sw P35434: ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR (EC 3.6.1.34)." "Mus musculus F1FOATP synthase complex E subunit (Atb5k) gene, complete cds"  
"Mus musculus NAD(H)-specific isocitrate dehydrogenase gamma subunit precursor, mRNA, complete  
"M.musculus gene for dodecenoyl-CoA delta-isomerase, exons 1 and 2"  
"Mus musculus cytochrome c oxidase subunit VIII-H precursor (COX8H) mRNA, complete cds"  
"Homologous to sw P35745: ACYLPHOSPHATASE, MUSCLE TYPE ISOZYME (EC 3.6.1.7) (ACYLPHOSPHATE PH  
"Mus musculus CD-1 cardiac troponin I mRNA, complete cds"  
"Homologous to sw P00566: CREATINE KINASE, M CHAIN (EC 2.7.3.2) (NU-2 PROTEIN)." "Mouse mRNA for protein with homology to transition protein 2 (TP2)  
Mus musculus Selenium-binding liver protein mRNA  
"Mus musculus (clone MAR1) aldose reductase mRNA, complete cds"  
"Mus musculus vascular endothelial growth factor B 186 (VEGF-B) precursor, mRNA, complete cds"  
M.musculus mRNA for NADH transhydrogenase  
"Mus musculus aldehyde dehydrogenase (ALDH2) mRNA, nuclear gene encoding mitochondrial protein  
"Mouse cytosolic epoxide hydrolase mRNA, complete cds"  
"Mus musculus 129SV carnitine palmitoyltransferase II mRNA, complete cds"  
"Mus musculus medium-chain acyl-CoA dehydrogenase mRNA, complete cds"  
"Mus musculus long-chain acyl-CoA dehydrogenase mRNA, complete cds"  
"Mus musculus very-long chain acyl-CoA dehydrogenase mRNA, partial cds"  
"Mouse muscle creatine kinase mRNA (EC 2.7.3.2)  
"Mus musculus isocitrate dehydrogenase mRNA, complete cds"  
"Mus musculus long chain fatty acyl CoA synthetase mRNA, complete cds"  
"Mus musculus sterol carrier protein-2 (SCP-2) gene, complete cds"  
"Mouse alpha-tubulin isotype M-alpha-4 mRNA, complete cds"

# Cardiomyopathy: Downregulated genes



## BLAST Definitions

"Homologous to sm 007021: PRE-MRNA SPLICING FACTOR SF2. P32 SUBUNIT PRECURSOR (GCTQ-R PROTEIN)  
"Mus musculus proteasome activator PA28 alpha subunit mRNA, complete cds"  
"Mus musculus cdc37 homolog mRNA, complete cds"  
"Mus musculus ornithine decarboxylase antizyme gene, complete cds"  
M.musculus mRNA for carnitine acetyltransferase  
"Homologous to sm 000779: CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE (EC 3.6.1.34)  
"Homologous to sm P11507: CALCIUM-TRANSPORTING ATPASE ENDOPLASMIC RETICULUM TYPE (EC 3.6.1.38)  
M.musculus ENO3 mRNA for enolase beta subunit  
"Homologous to sm P47858: 6-PHOSPHOFRUCTOKINASE. MUSCLE TYPE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE)  
"Homologous to sm P23327: SARCOPLASMIC RETICULUM HISTIDINE-RICH CALCIUM-BINDING PROTEIN PRECURS  
"Mouse AE3 mRNA, complete cds"  
"M.musculus glucose transporter 2 mRNA, complete cds"  
Mus musculus aspartate aminotransferase gene 5'-flank and exon 1  
"Mus musculus thioredoxin-dependent peroxide reductase (txn) mRNA, complete cds"  
"Homologous to sm P47858: 6-PHOSPHOFRUCTOKINASE. MUSCLE TYPE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE)  
"Homologous to sm P11508: CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE (EC 3.6.1.38)  
"Homologous to sm P35434: ATP SYNTHASE DELTA CHAIN. MITOCHONDRIAL PRECURSOR (EC 3.6.1.34)."  
"Mus musculus F1FOATP synthase complex E subunit (Atm5k) gene, complete cds"  
"Mus musculus NAD(H)-specific isocitrate dehydrogenase gamma subunit precursor, mRNA, complete  
"M.musculus gene for dodecenoyl-CoA delta-isomerase, exons 1 and 2"  
"Mus musculus cytochrome c oxidase subunit VIII-H precursor (COX8H) mRNA, complete cds"  
"Homologous to sm P35745: ACYLPHOSPHATASE. MUSCLE TYPE ISOZYME (EC 3.6.1.7) (ACYLPHOSPHATE PHOSPHATASE)  
"Mus musculus CD-1 cardiac troponin I mRNA, complete cds"  
"Homologous to sm P00566: CREATINE KINASE. M CHAIN (EC 2.7.3.2) (NU-2 PROTEIN)."  
"Mouse mRNA for protein with homology to transition protein 2 (TP2)  
Mus musculus Selenium-binding liver protein mRNA  
"Mus musculus (clone MAR1) aldose reductase mRNA, complete cds"  
"Mus musculus vascular endothelial growth factor B 186 (VEGF-B) precursor, mRNA, complete cds"  
M.musculus mRNA for NADH transhydrogenase  
"Mus musculus aldehyde dehydrogenase (ALDH2) mRNA, nuclear gene encoding mitochondrial protein  
"Mouse cytosolic epoxide hydrolase mRNA, complete cds"  
"Mus musculus 129SV carnitine palmitoyltransferase II mRNA, complete cds"  
"Mus musculus medium-chain acyl-CoA dehydrogenase mRNA, complete cds"  
"Mus musculus long-chain acyl-CoA dehydrogenase mRNA, complete cds"  
"Mus musculus very-long chain acyl-CoA dehydrogenase, partial cds"  
"Mouse muscle creatine kinase mRNA (EC 2.7.3.2)  
"Mus musculus isocitrate dehydrogenase mRNA, complete cds"  
"Mus musculus long chain fatty acyl CoA synthetase mRNA, complete cds"  
"Mus musculus sterol carrier protein-2 (SCP-2) gene, complete cds"  
"Mouse alpha-tubulin isotype M-alpha-4 mRNA, complete cds"

Fatty Acid Degradation?  
Other pathways / processes?

# What do we really need? Well...



## Find the pathways:

Biological processes in duodenal mucosa affected by glutamine administration

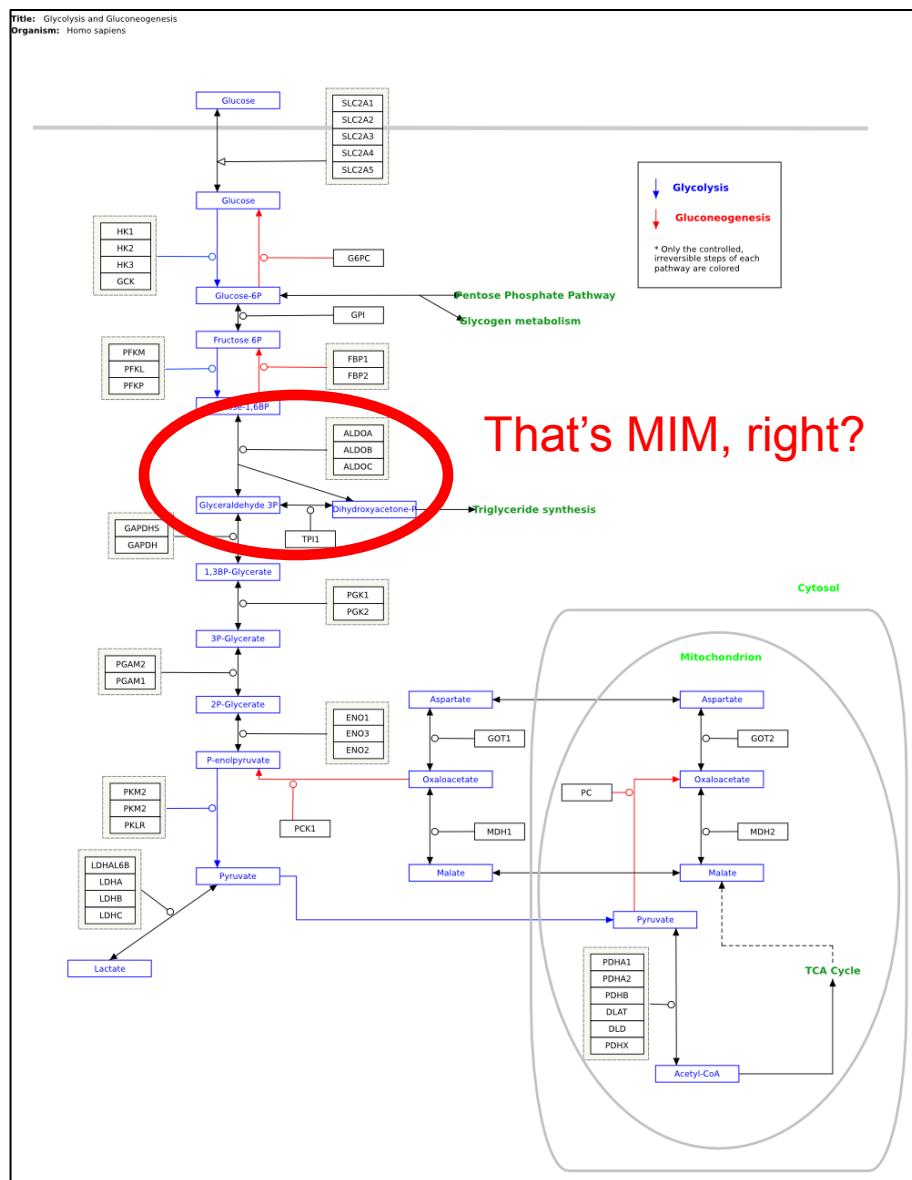
Pathway	number of genes					
	Changed	Up	Down	Measured	Total	Z Score
<a href="#">Hs_Mitochondrial_fatty_acid_betaoxidation</a>	6	6	0	16	16	4.456
<a href="#">Hs_Electron_Transport_Chain</a>	17	17	0	85	105	4.278
<a href="#">Hs_Fatty_Acid_Synthesis</a>	5	5	0	21	22	2.757
<a href="#">Hs_Fatty_Acid_Beta-Oxidation</a>	6	6	0	31	32	2.424
<a href="#">Hs_mRNA_processingReactome</a>	16	6	10	118	127	2.402
<a href="#">Hs_Unsaturated_Fatty_Acid_Beta_Oxidation</a>	2	2	0	6	6	2.342
<a href="#">Hs_HSP70_and_Apoptosis</a>	4	4	0	18	18	2.299
<a href="#">Hs_Oxidative_Stress</a>	5	5	0	27	28	2.097
<a href="#">Hs_Fatty_Acid_Omega_Oxidation</a>	3	3	0	14	15	1.915
<a href="#">Hs_Proteasome_Degradation</a>	8	8	0	60	61	1.629
<a href="#">Hs_RNA_transcriptionReactome</a>	5	5	0	38	40	1.25
<a href="#">Hs_Irinotecan_pathway_PharmGKB</a>	2	1	1	12	12	1.154
<a href="#">Hs_Synthesis_and_Degradation_of_Ketone_BodiesKEGG</a>	1	1	0	5	5	1.023

# Understanding genomics

## Example WikiPathways Pathway

Pathway on glycolysis.  
Using modern systems  
biology annotation.

And genes and  
metabolites connected to  
major databases.



PathVisio / WikiPathways - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Most Visited Getting Started Latest Headlines

PathVisio / WikiPathways

PathVisio and WikiPathways Development

Wiki Timeline Roadmap Browse Source View Tickets Search

Login | Help/Guide | About Trac | Preferences | Register

Start Page Index History Last Change

More... IUC Pathway:Pathway.gnml - PathVisio 2

Draw for presentation and publication

Visualize microarray or metabolomics data

Link to genomic databases and literature

Share pathway data on wikiPathways.org

Download Now Visual Tour

News

Aug 27, 2010 PathVisio 2.0.8 Has been released! See [ReleaseNotes](#)

**PathVisio / WikiPathways Development**

This is the wiki + bug tracker for [PathVisio](#) and [WikiPathways](#).

**License**

PathVisio is licensed to you under the [Apache License, version 2.0](#).

**Attachments**

- PathVisioSite-1.png (135.8 KB) - added by [martijn](#) 11 months ago.

Done

# PathVisio

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

- Data modeling and visualisation on biological pathways
- It can use gene expression, proteomics and metabolomics data
- Identify significantly changed processes

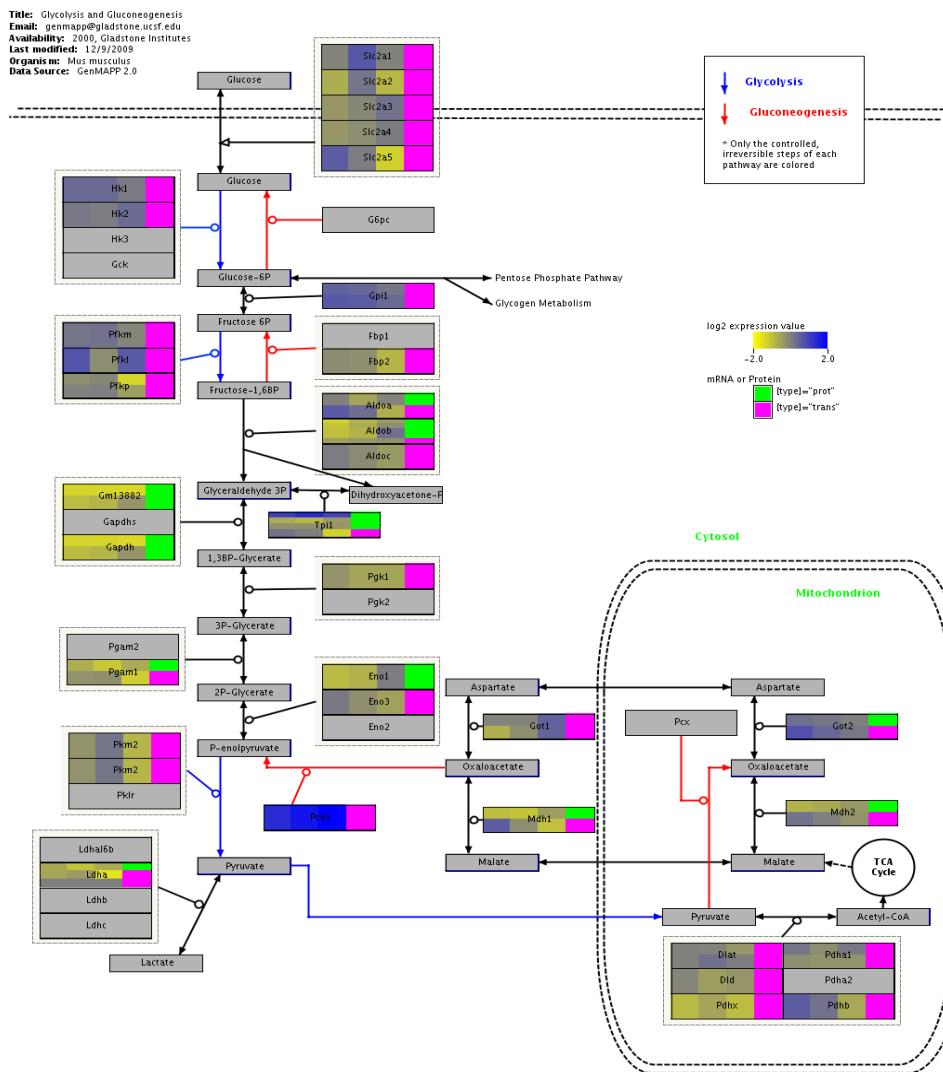
Martijn P van Iersel, Thomas Kelder, Alexander R Pico, Kristina Hanspers, Susan Coort, Bruce R Conklin, Chris Evelo (2008) Presenting and exploring biological pathways with PathVisio. BMC Bioinformatics 9: 399

# adding data = adding colour

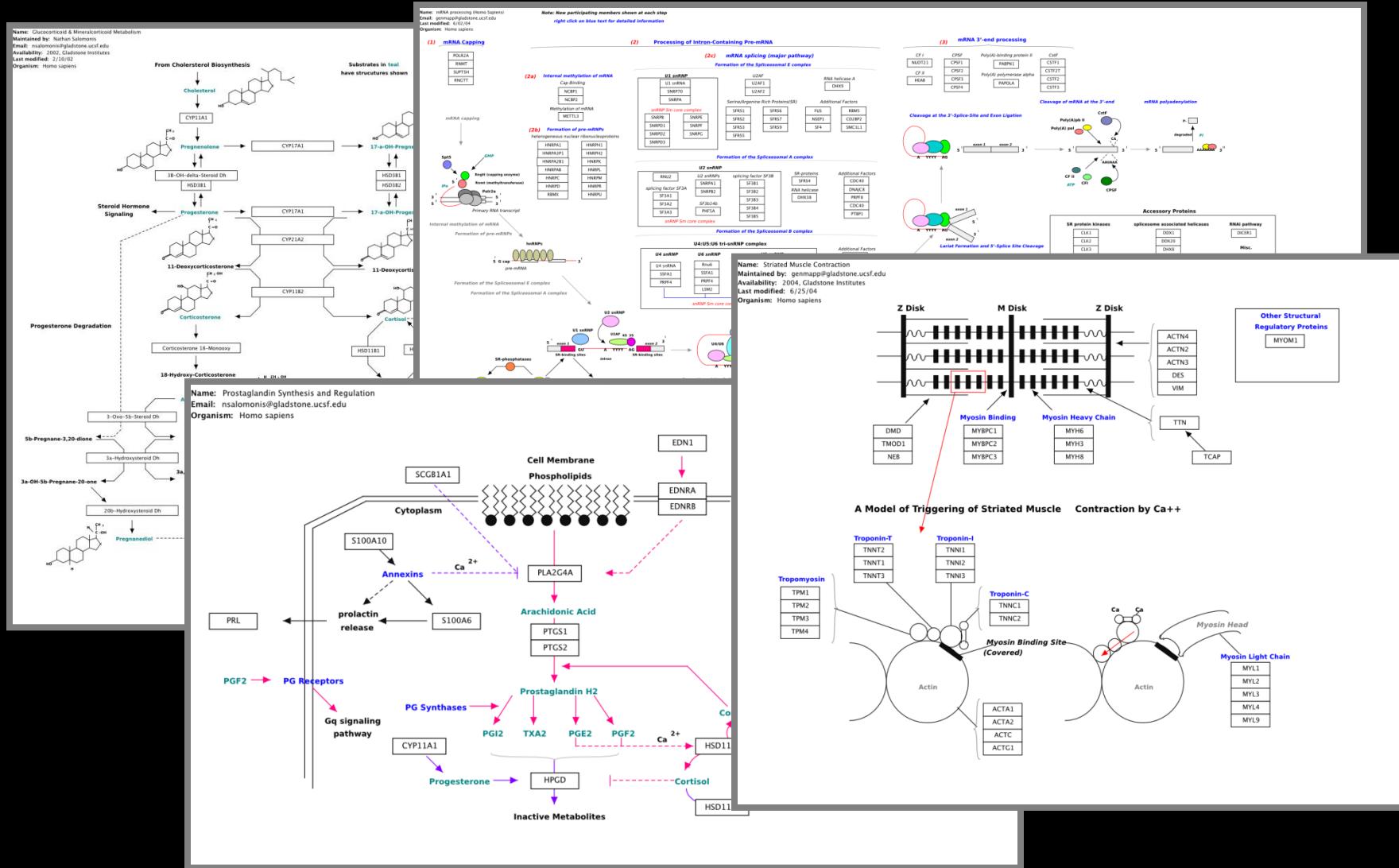
## Example PathVisio result

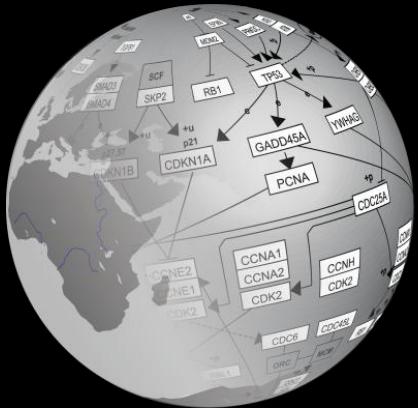
Showing proteomics and transcriptomics results on the glycolysis pathway in mice liver after starvation.

[Data from Kaatje Lenaerts and Milka Sokolovic, analysis by Martijn van Iersel]



# Now we just need the Pathways





# WikiPathways

- Public resource for biological pathways
- **Anyone** can contribute and curate
- More up-to-date representation of biological knowledge

**WikiPathways: Pathway Editing for the People.** Alexander R. Pico, Thomas Kelder, Martijn P. van Iersel, Kristina Hanspers, Bruce R. Conklin, Chris Evelo. **PLoS Biology** 2008; 6: 7. e184

Commentaries:

**Big data: Wikiomics.** Mitch Waldrop. **Nature** 2008; 455, 22-25

**We the curators.** Allison Doerr. **Nature Methods** 2008; 5, 754–755

**No rest for the bio-wikis.** Ewen Callaway. **Nature** 2010; 468, 359-360

# How to do we do data visualization?

**Ce\_RNA\_interference\_and\_miRNA39386 - GenMAPP 2.1**

Clipboard Font Alignment Styles Format Filter Select

	A	B	D	E	
1	ID	System	t0	t1	t2
2	177140_s_X		0.693	-2.856	
3	186328_at_X		1.327	3.497	
4	192151_at_X		3.863	-3.457	
5	187703_s_X		1.16	-3.352	
6	193913_s_X		-0.435	3.356	
7	173437_at_X		-2.36	-1.633	
8	173452_s_X		-2.43	0.035	
9	no affy	X	-1.878	2.211	
10	178221_at_X		-1.758	-3.223	
11	173832_at_X		-3.386	1.695	
12	184510_at_X		0.109	-1.045	
13	192371_s_X		-3.075	-0.404	
14	172670_x_X		3.642	3.428	
15	176778_s_X		-0.112	1.072	-1.695

**RNA interference and miRNA**

dsRNA

Dicer (Dcr-1)

primary miRNA

miRNA processor complex

DRSH-1 PASH-1

RISC Complex

RRF1, RRF2, EGO1, RRF3

inhibitor / competitor

RNA dependent RNA polymerase family

argonaute family: RDE-1, PPW-1; VIG-1, TSN-1

Silencing

Some famous miRNAs: let-7, lin-41; lin-4, lin-14, lin-28

Other genes involved in RNAi: RDE-2/MUT-8, MUT-7, RDE-3/MUT-2, RDE-4

Legend: t0

- Up
- Down
- No criteria met
- Not found

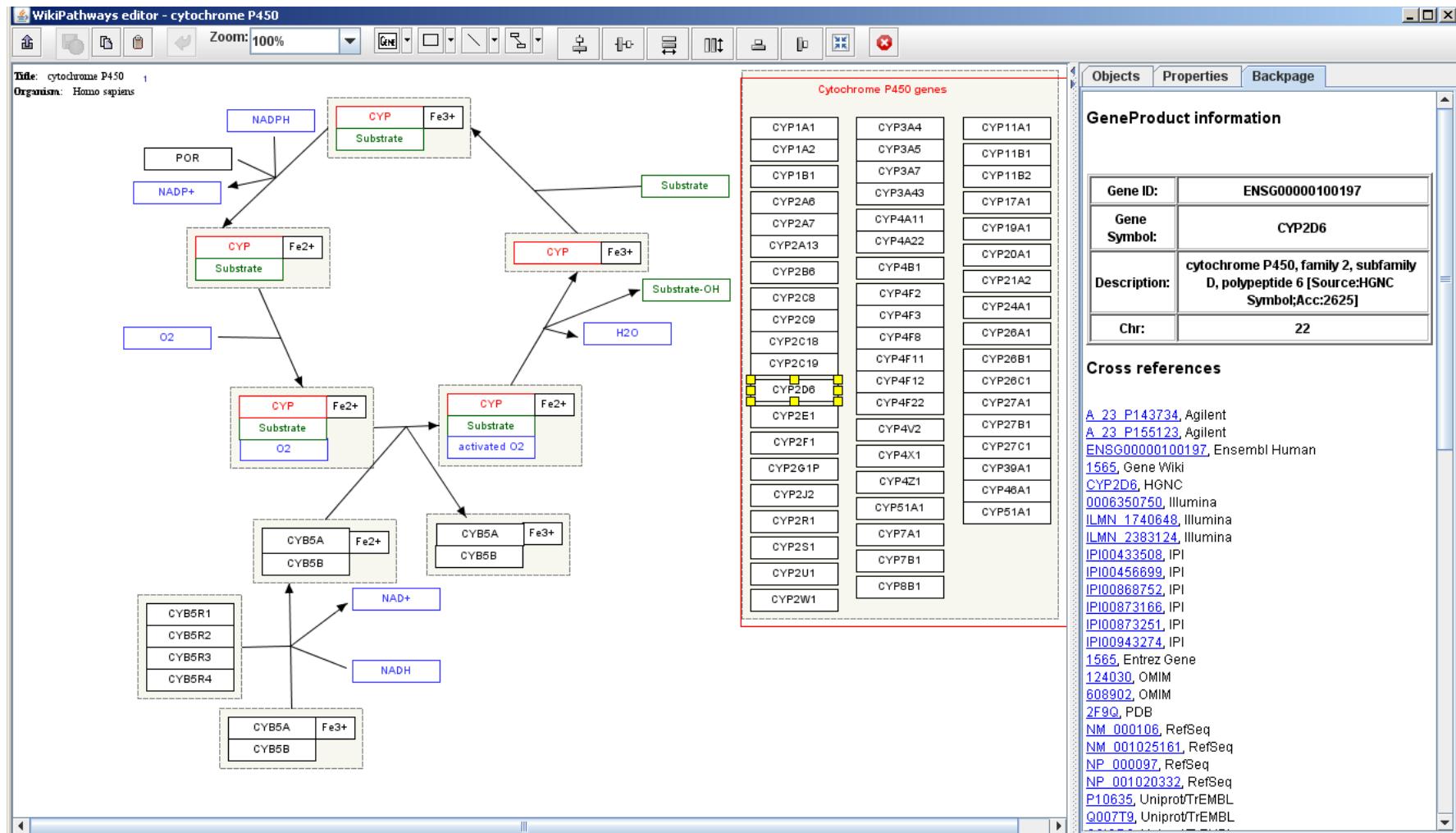
Author: Martijn van Iersel  
Maintained by: Martijn van Iersel  
E-mail: martijn.vaniersel@bigcat.unimaas.nl

# Connect to Genome Databases

The image displays three distinct software environments related to genome databases:

- GenMAPP 2.0:** A desktop application window titled "Ce\_RNA\_interference\_and\_miRNA - GenMAPP 2.0". It features a menu bar (File, Tools, Format, View, Data, Help) and a toolbar with various icons. The main area contains a diagram illustrating the RNA interference pathway. It shows double-stranded RNA (dsRNA) being processed by Dicer (DCR-1) to produce 22nt siRNA. This siRNA then triggers a feedback loop labeled "Feedback / Spreading" through an RNA-dependent RNA polymerase family (RRF-1, RRF-2, EGO-1, RRF-3). An inhibitor/competitor (argin) is also shown. The diagram includes labels like "inhibitor / competitor" and "Feedback / Spreading".
- Web Browser:** A standard web browser window showing the WormBase website. The URL is http://www.wormbase.org/db/gene/gene?name=WBGene00006924. The page title is "RNA interference and miRNA". The header includes "Some famous miRNA's". The main content area shows a search bar with "Find: WBGene00006924" and a dropdown menu set to "Anything". Below the search bar are links for "Home", "Genome", "Blast / Blat", "WormMart", "Batch Sequences", "Markers", "Genetic Maps", "Submit", "Searches", and "Site Map". The WormBase logo is prominently displayed.
- Gene Summary Page:** A detailed gene summary page for the gene *vig-1*. The page title is "Gene Summary for vig-1". It prompts the user to specify a gene using a gene name (*unc-26*), a predicted gene id (R13A5.9), or a protein ID (CE02711). Below this is a link to "[identification][location][function][expression][gene ontology][alleles][similarities][reagents][bibliography]". The page is divided into several sections:
  - Identification:** Shows the main name as *vig-1* - (VIG (Drosophila Vasa Intronic Gene) ortholog) via person evidence: Ronald Plasterk. It also lists the Sequence name as F56D12.5 and the Other name(s) as 2B613 (inferred automatically). The WB Gene ID is WBGene00006924.
  - Concise Description:** States that *vig-1* encodes a predicted RNA-binding protein orthologous to Drosophila VIG (Vasa Intronic Gene); *vig-1* activity is required for proper function of the *let-7* miRNA in vivo and thus, for regulating the transition from late larval to adult cell fates; VIG-1 is a component of the 250 kDa RNA-induced silencing complex (RISC) complex and co-immunoprecipitates with both TSN-1, the *C. elegans* Tudor-SN ortholog, and the *let-7* miRNA. [details]
  - NCBI KOGs\*:** Predicted RNA-binding protein [KOG2945]; [OMpre\_WH000977]
  - Species:** *Caenorhabditis elegans*
  - NCBI:** [AceView: 2B613]

# Backpages link to databases



# BridgeDB: Abstraction Layer

**interface**  
IDMapper

**class**  
IDMapperRdb  
*relational database*

**class**  
IDMapperFile  
*tab-delimited text*

**class**  
IDMapperBiomart  
*web service*



BridgeDB Identifier Mapping Uses:

MIRIAM and Identifiers.org

to provide Java framework

# Download Pathways

p38 MAPK Signaling Pathway (Homo sapiens) - WikiPathways - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Open PHACTS: Sign in

www.wikipathways.org/index.php/Pathway:WP400#nogo2

http://www.biogcat.un...

Most Visited Getting Started Latest Headlines Import to Mendeley

Evelo my talk my preferences my watchlist my contributions log out

pathway discussion watch

BETA WIKIPATHWAYS Pathways for the People

search

navigation

- Home
- Help

pathway

- Create
- Browse
- Wish List
- Download
- Web service API

overview

- Recent Changes
- Most Viewed
- Most Edited
- New Pathways
- Statistics

community

- About us
- Contact us
- How to cite
- Curation events
- BiGCaT portal
- CIRM portal
- GenMAPP portal
- Micronutrient portal
- NetPath portal
- Reactome portal
- Development

toolbox

- What links here
- Related changes
- Upload file

pathway

watch

## p38 MAPK Signaling Pathway (Homo sapiens)

Kristina Hanspers, Thomas Kelder, ramesh , Alexander Pico, et al.

Title: p38 MAPK Signaling Pathway Availability: CC BY 2.0 Organism: Homo sapiens

Diagram illustrating the p38 MAPK Signaling Pathway. Key components include SREBP1, ERK1/2, JNK1/2, and p38 MAPK. The pathway shows various stimuli (e.g., Growth Factors, UV, Stress, Oncogenic, etc.) activating these kinases through cascades involving MAP3K1, MAP3K2, and MAP3K3. These kinases then activate MAPK14, which leads to various cellular responses such as Transcription, Apoptosis, and DNA transcription. The diagram also shows interactions with other pathways like PI3K-Akt and NF-κB.

Contents

- 1 Curation
- Tags
- 2 Description
- 3 Ontology
- 4 Bibliography
- 5 Categories
- 6 History
- 7 External references

Edit pathway not working?

Download

- PathVisio (.gpmi)
- Scalable Vector Graphics (.svg)
- Gene list (.bt)
- Biopax level 3 (.owl)
- Eu.Gene (.pwt)
- Png image (.png)
- Acrobat (.pdf)
- GenMAPP (.mapp)

Curation Tags

hide

Featured pathway

http://www.wikipathways.org/index.php/Pathway:WP400#nogo2

# GPML Cytoscape Plugin

[http://www.pathvisio.org/wiki/Cytoscape\\_plugin](http://www.pathvisio.org/wiki/Cytoscape_plugin)

The GPML plugin for Cytoscape is a converter between Cytoscape networks and the GPML (GenMAPP Pathway Markup Language) pathway format. This plugin makes it possible to use GPML pathways (which you can find on [WikiPathways](#)) as Cytoscape networks and improve the GPML pathway content using Cytoscape features.

The p38 MAPK signalling pathway in Cytoscape

Extending GPML pathway using the [LitSearch?](#) plugin

Example pathway in Cytoscape (left) and [PathVisio](#) (right)

Finding the shortest path between two nodes on a GPML pathway

## Features

- Open [PathVisio](#) pathways in Cytoscape, including visual annotations
- Save Cytoscape networks to the GPML format, for use in [PathVisio](#)
- Copy/paste pathway elements or nodes and edges between Cytoscape and [PathVisio](#)
- Search and download pathways from [WikiPathways](#)

You can install the plugin via the plugin manager in Cytoscape or download it [here](#).

You can also view and download the latest source code from the [svn repository](#).

**Note:** By default, the plugin only converts GPML elements of type `DataNode?` to a cytoscape node. To convert Label elements as well, set the property "gpml.label.as.node" to "true" in Edit->Preferences->Properties.

## Applications

Improving pathways using Cytoscape::

The [LitSearch?](#) plugin is a great example of how Cytoscape can be used to improve pathway content. By using the 'extend network from literature' feature, you can easily find new genes and interactions that can be potential additions to the pathway, based on the genes that are already in there. Or you can build a new network based on a list of genes and export this network to GPML. This is much easier than building a pathway from scratch based on a pubmed search and literature study. The [LitSearch?](#) plugin could also be used to improve microarray analysis. For example, with MAPPFinder you often end up with a pathway that is just a list of genes (e.g. a GO category, or a pathway in the [Molecular Function category](#)), which would be much more informative if you would see interactions between these genes. If you copy this pathway to Cytoscape and apply the [LitSearch?](#) plugin, you have a network you can map the data onto (either in Cytoscape, or copied back into [PathVisio](#)/GenMAPP).

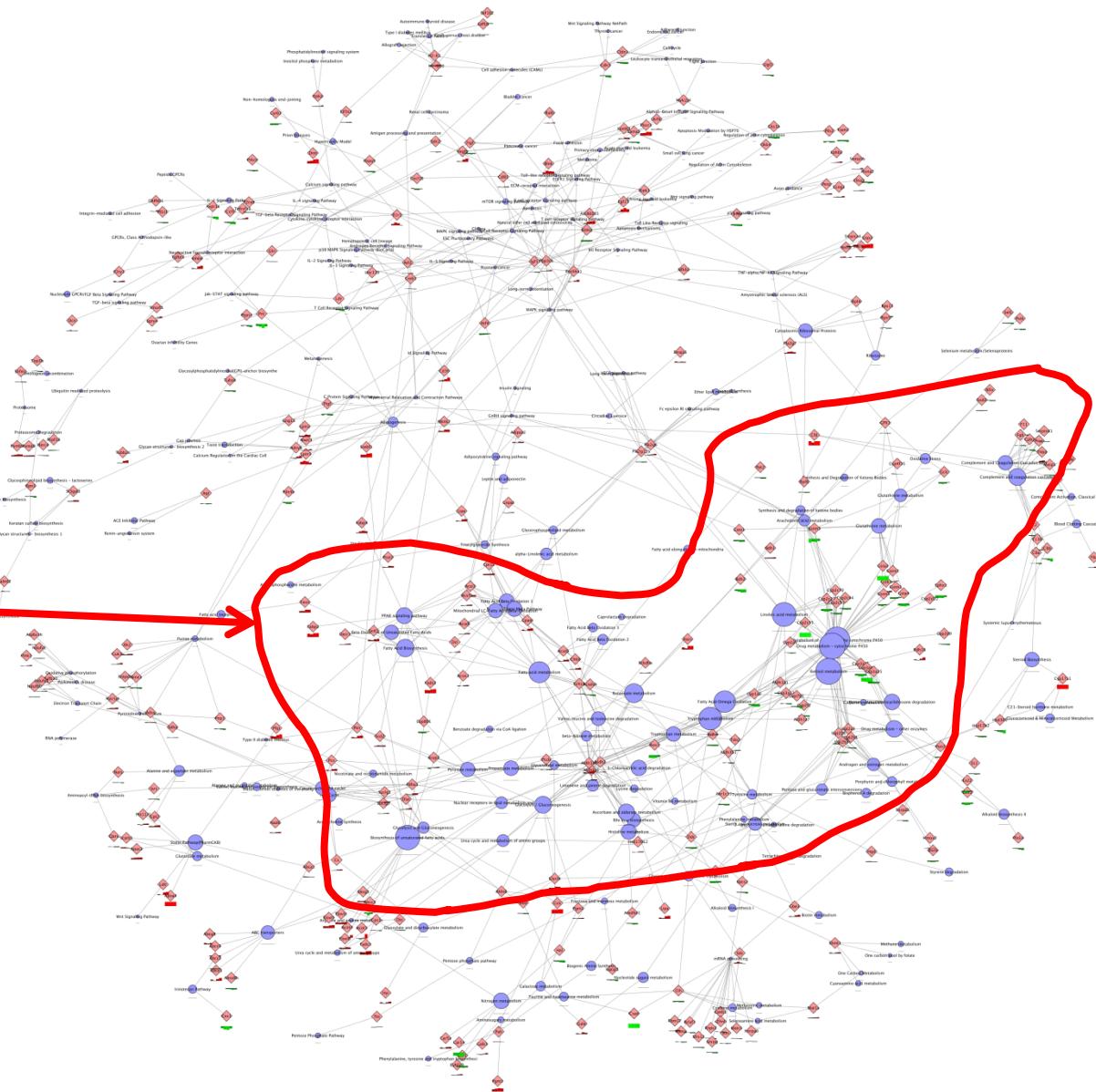
# Cytoscape visualization used to group

PPS1  
Liver

All pathways

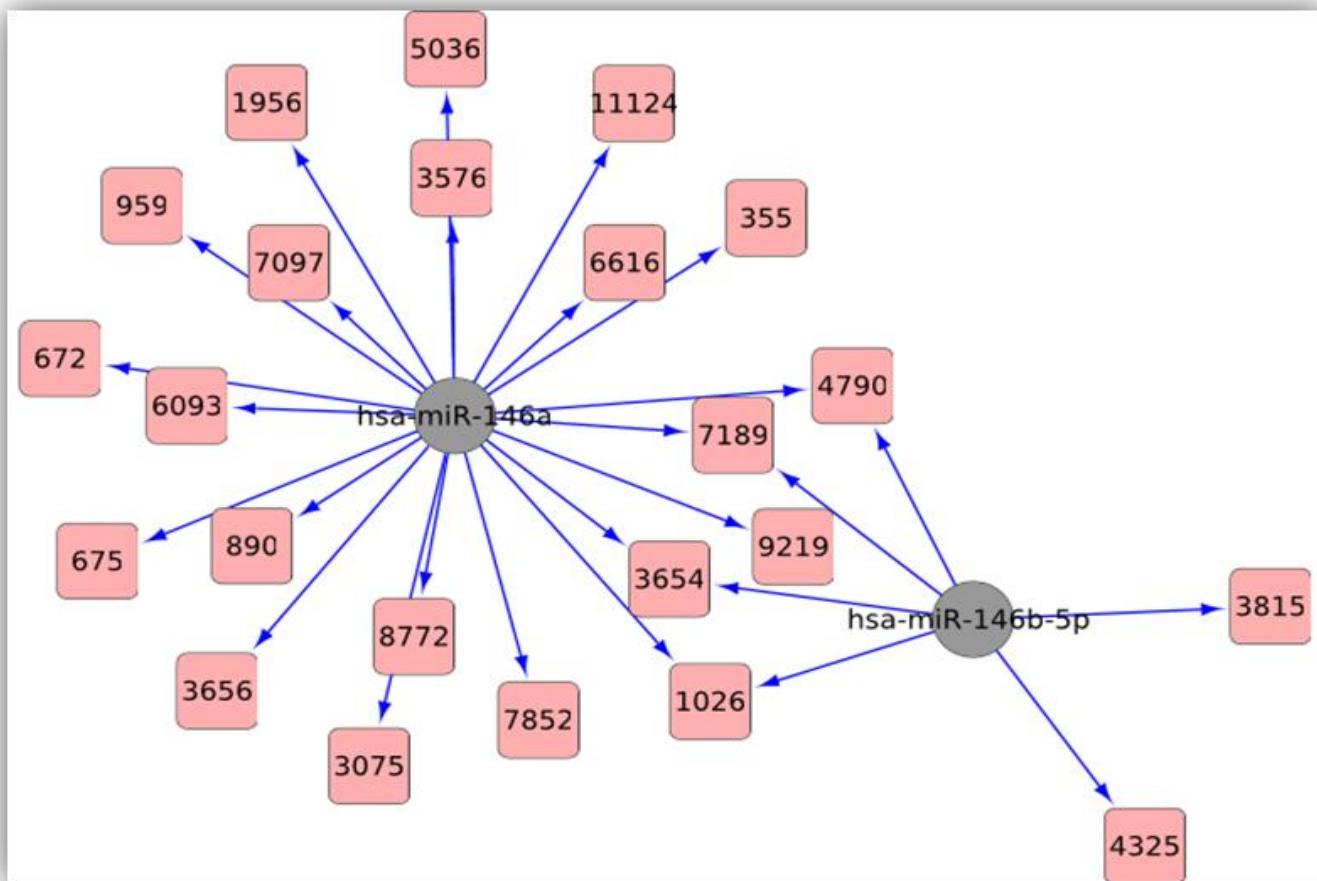
Pathways with high z-score grouped together.

Explains why there are relatively few significant genes, but many pathways with high z-score.

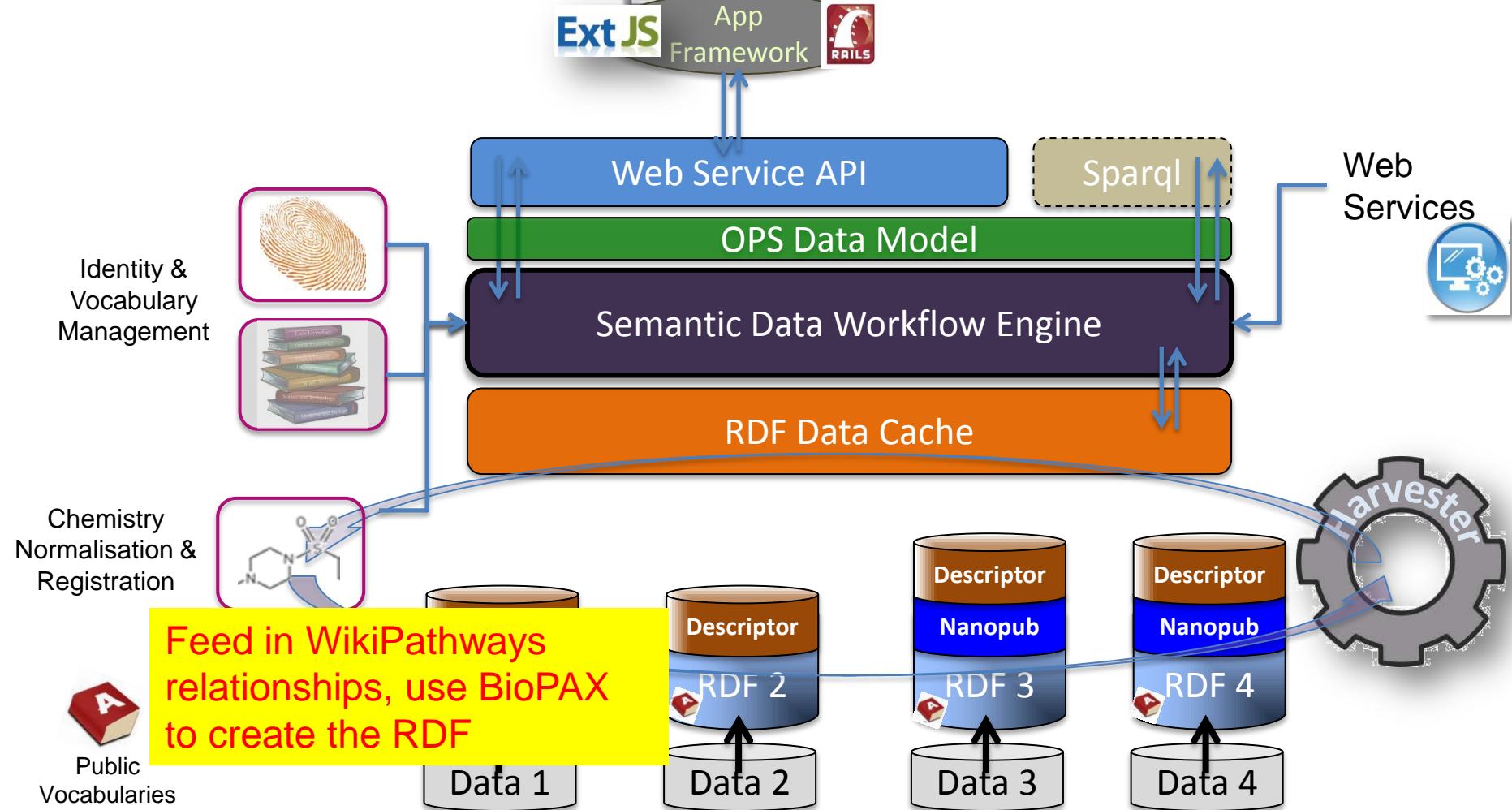


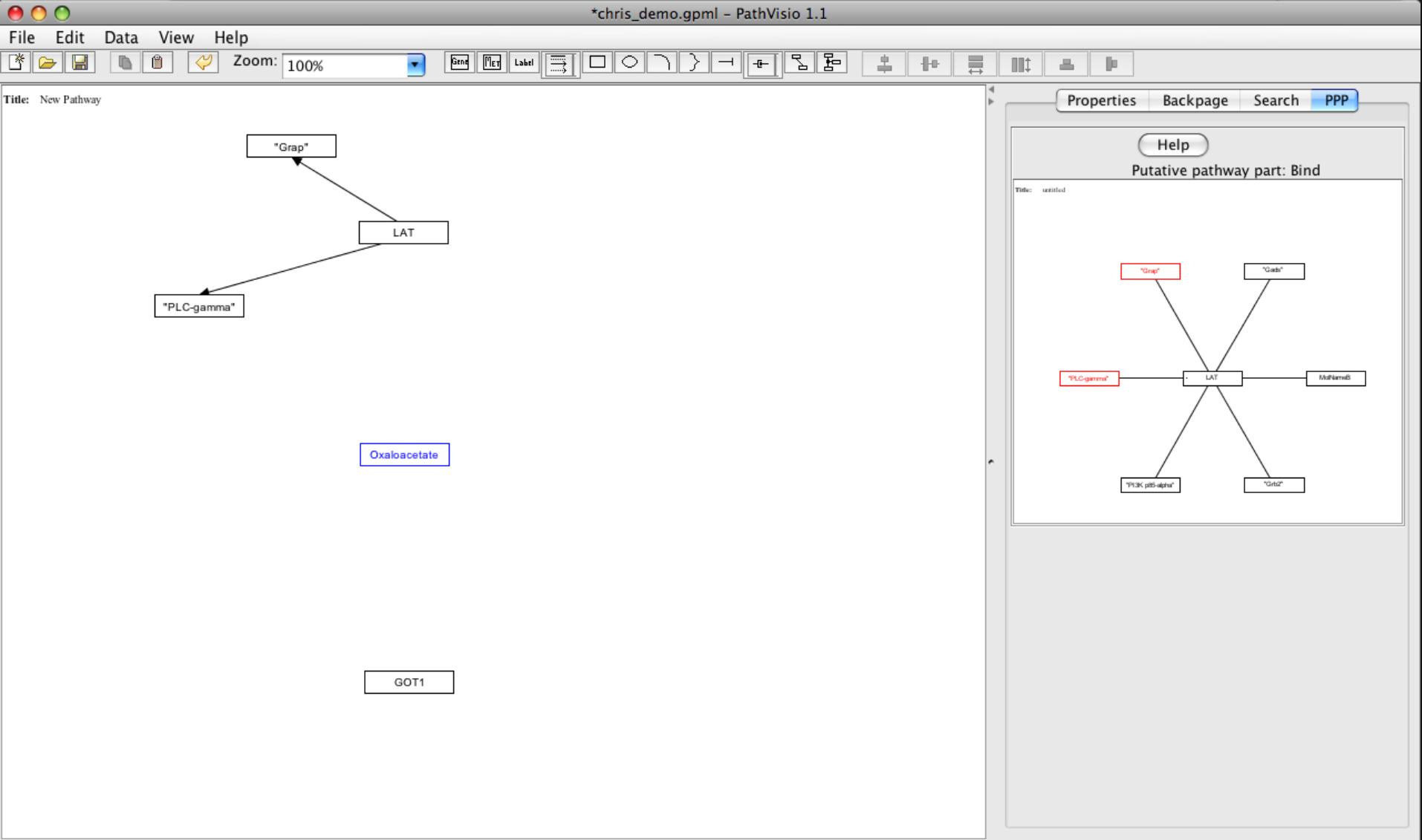
# MiRNAs of Interest

miRNA target information from miRBase

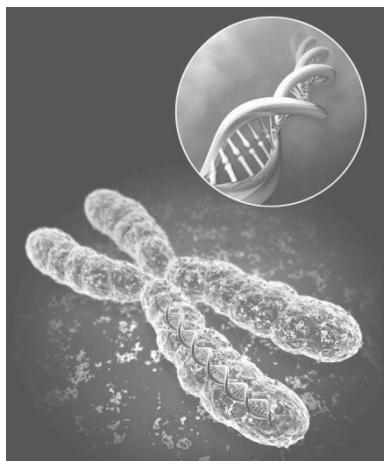
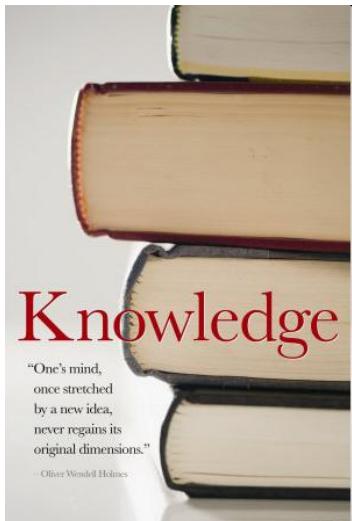


Note: Things may change!





| Gene database: .../Hs\_Derby\_20080102.pgdb | Metabolite database: ...metabolites\_080916.pgdb



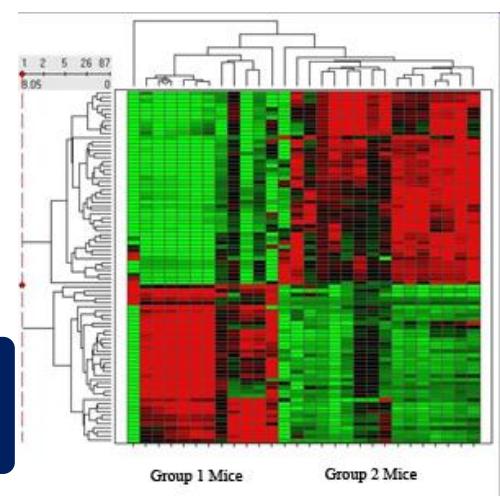
Existing  
knowledge

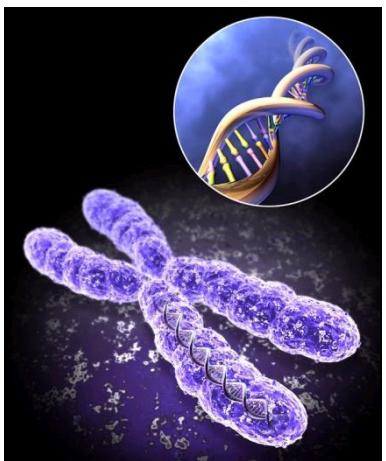
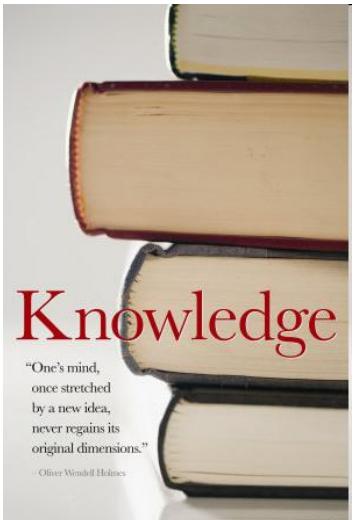


Genetic  
Results



Genomic  
Results





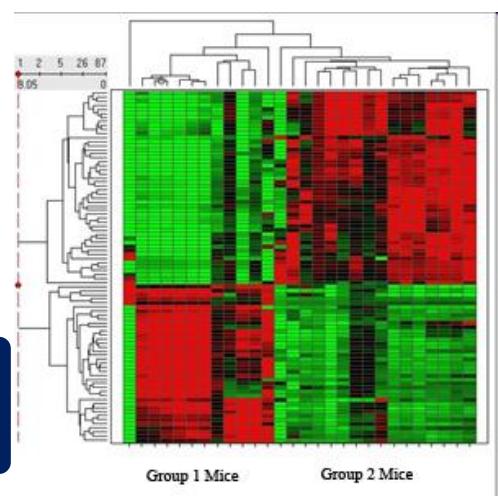
Existing  
knowledge

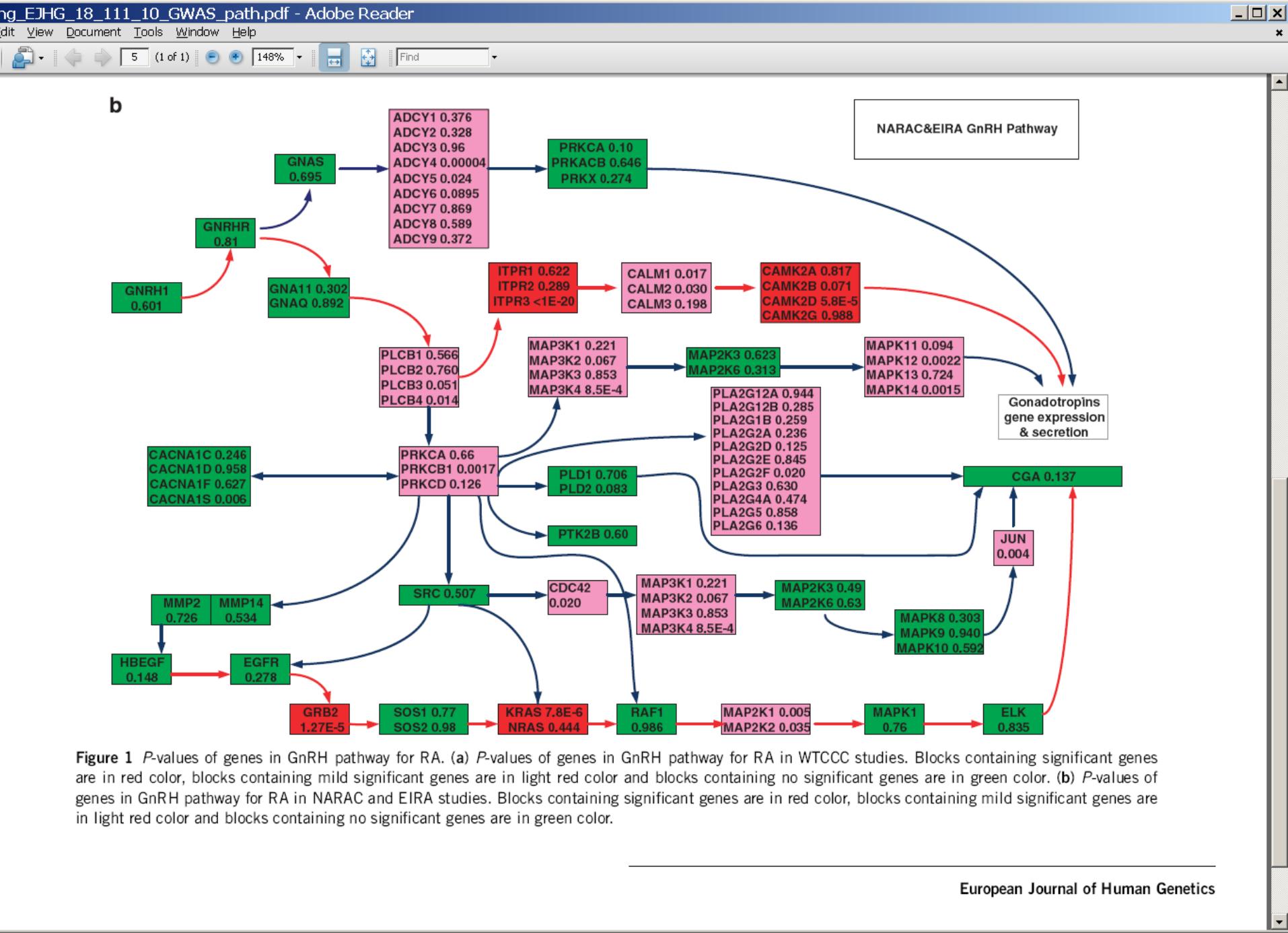


Genetic  
Results



Genomic  
Results

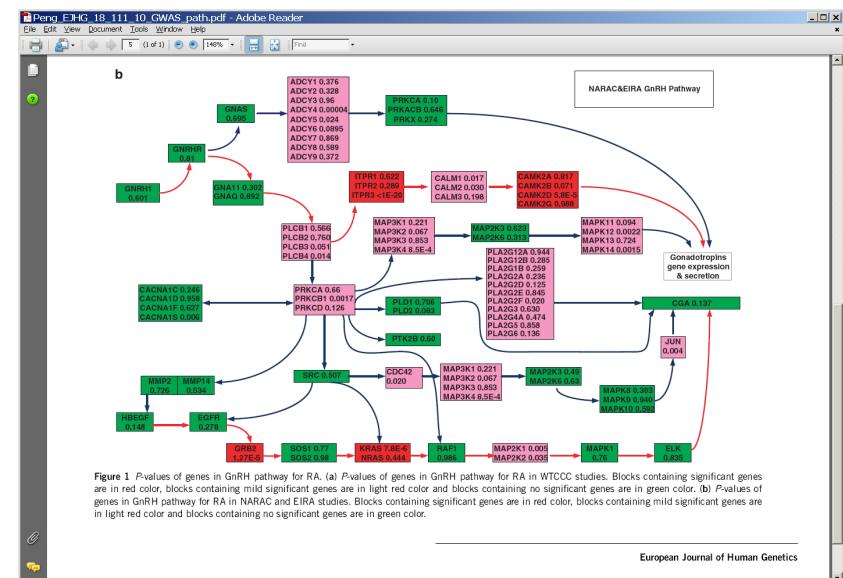




# Can we show SNPs?

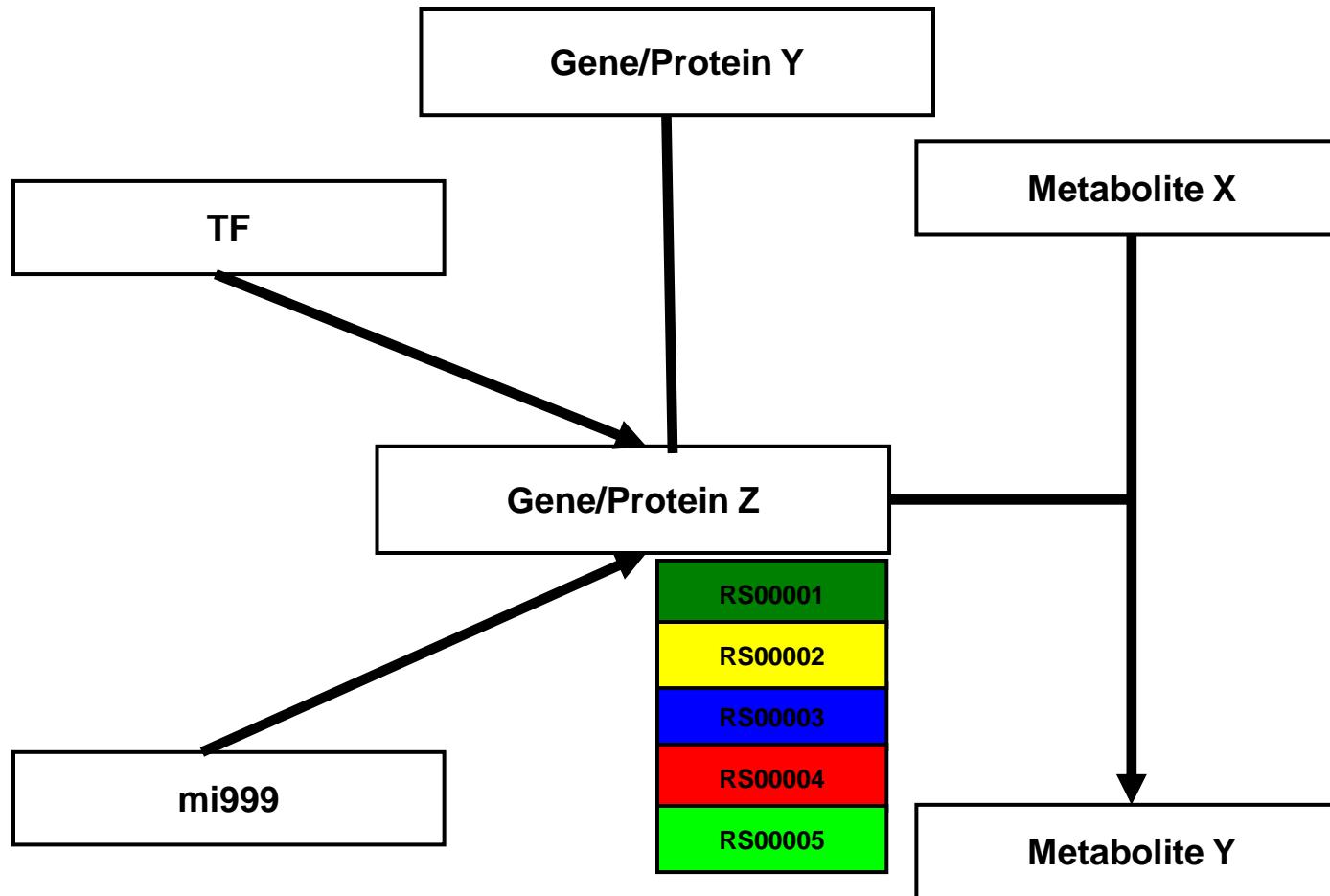
YES WE CAN

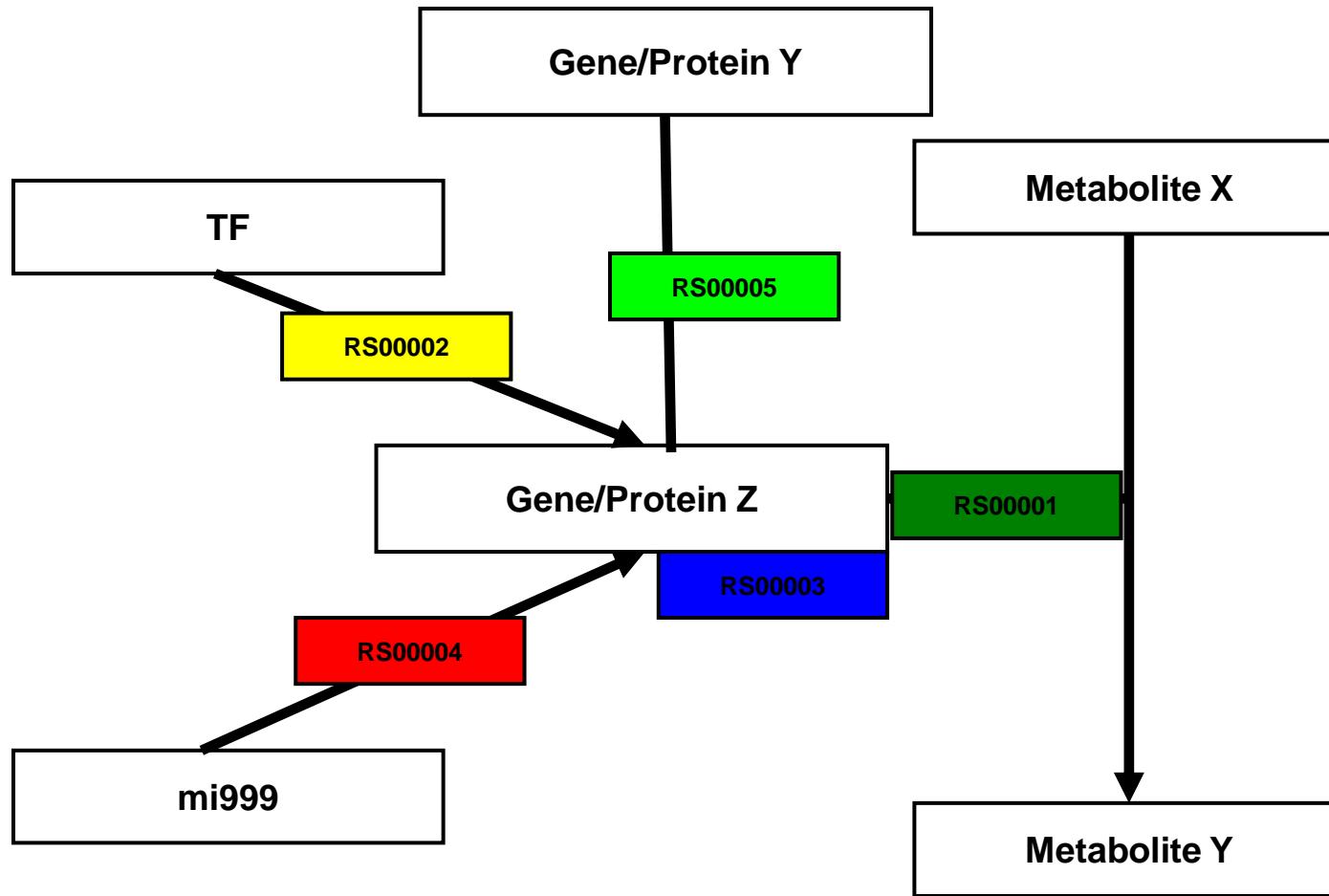
Using dbSNP links in ENSEMBL  
as part of BridgeDB libs



# But it will look like this....







## Functionalize SNPs

Unknown function (attribute to gene)

Changing protein functionality (coding)

In miRNA binding site

Changing protein interactions (coding)

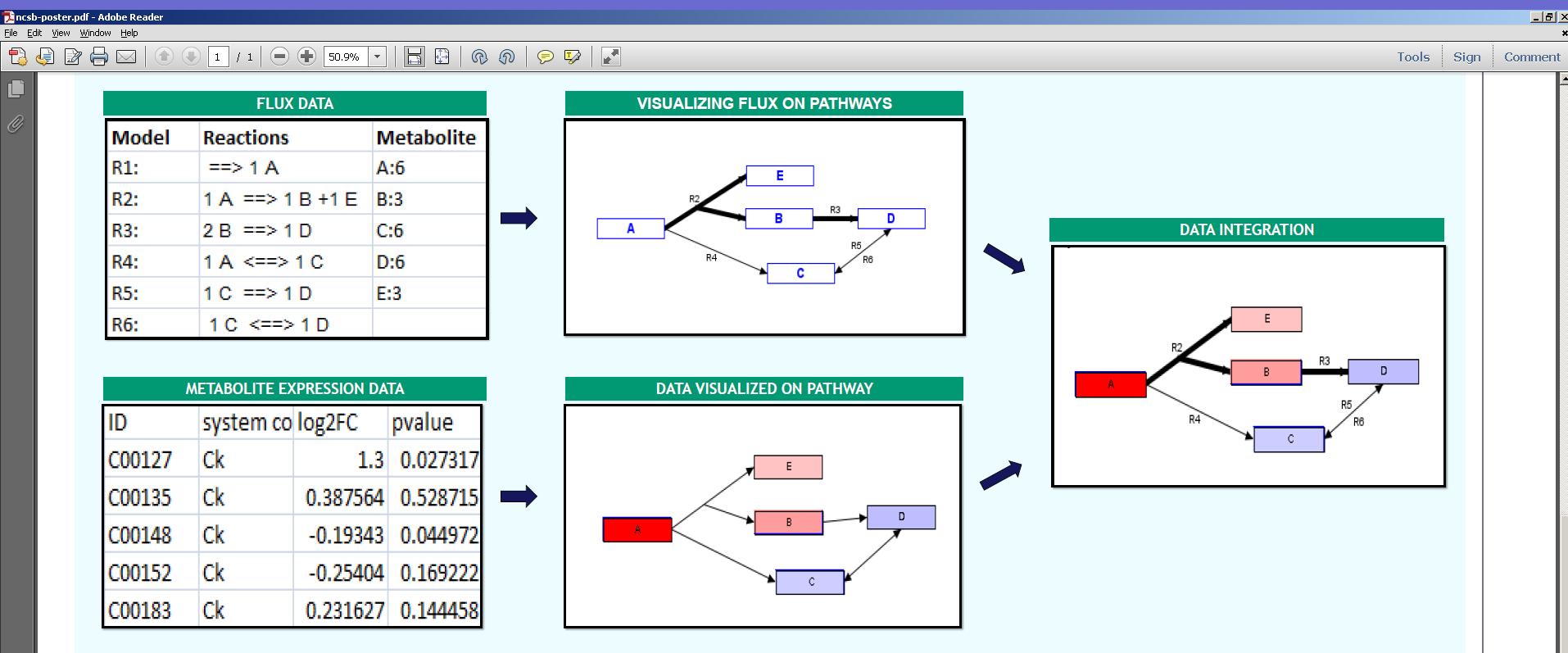
In TF binding site

# SNPs would affect the edges:

- Model the pathways for different SNPs
- Visualize the results (flux visualization plugin)
- Connect to actual modeling results (RepoSE)
- Allow integration of SBML - GPML

# Integrating it all

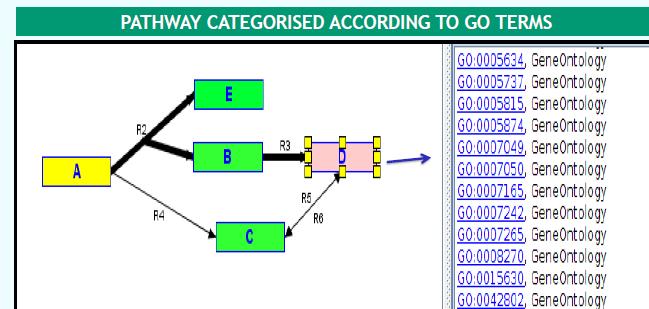
## Visualizing fluxes, data and annotation



### Project 2: GO Analysis Integration

- ❖ Categorization of pathways according to Gene Ontology terms
- ❖ Assignment of automatically inferred GO categories to all metabolites.
- ❖ GO visualization plugin to show gene categories (eg:cellular locations) on Pathways
- ❖ Building Metabolite Ontology( MO ) based on GO.

GO:0005813 Centrosome  
GO:0005737 Cytoplasm  
GO:0005634 :Cell Nucleus



# In summary we use/need:

- MIM and SBGN  
Pathway structuring and visualization
- BioPAX  
Pathway export/import  
As vocabulary to create RDF
- MIRIAM and Identifiers.org  
Identifier mapping
- Everything Cytoscape  
To treat pathways as network cores
- RepoSE  
To get modeling results for flux visualization
- SBML  
For model integration

# Thanks!

