WikiPathways: how open source code and open data can make omics technology more useful

@Chris_Evelo

Department of Bioinformatics - BiGCaT

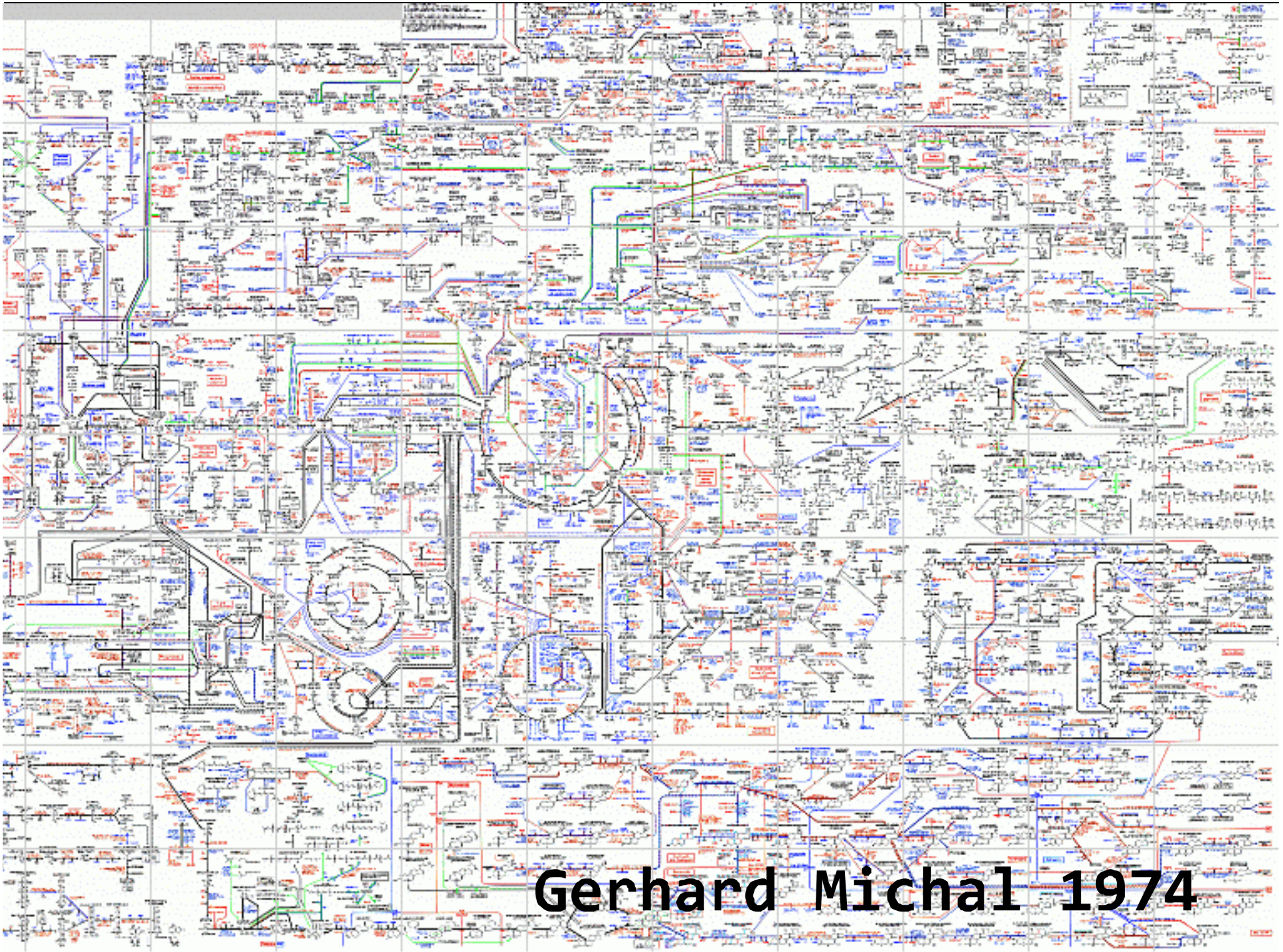
Maastricht University

www.BiGCaT.unimaas.nl

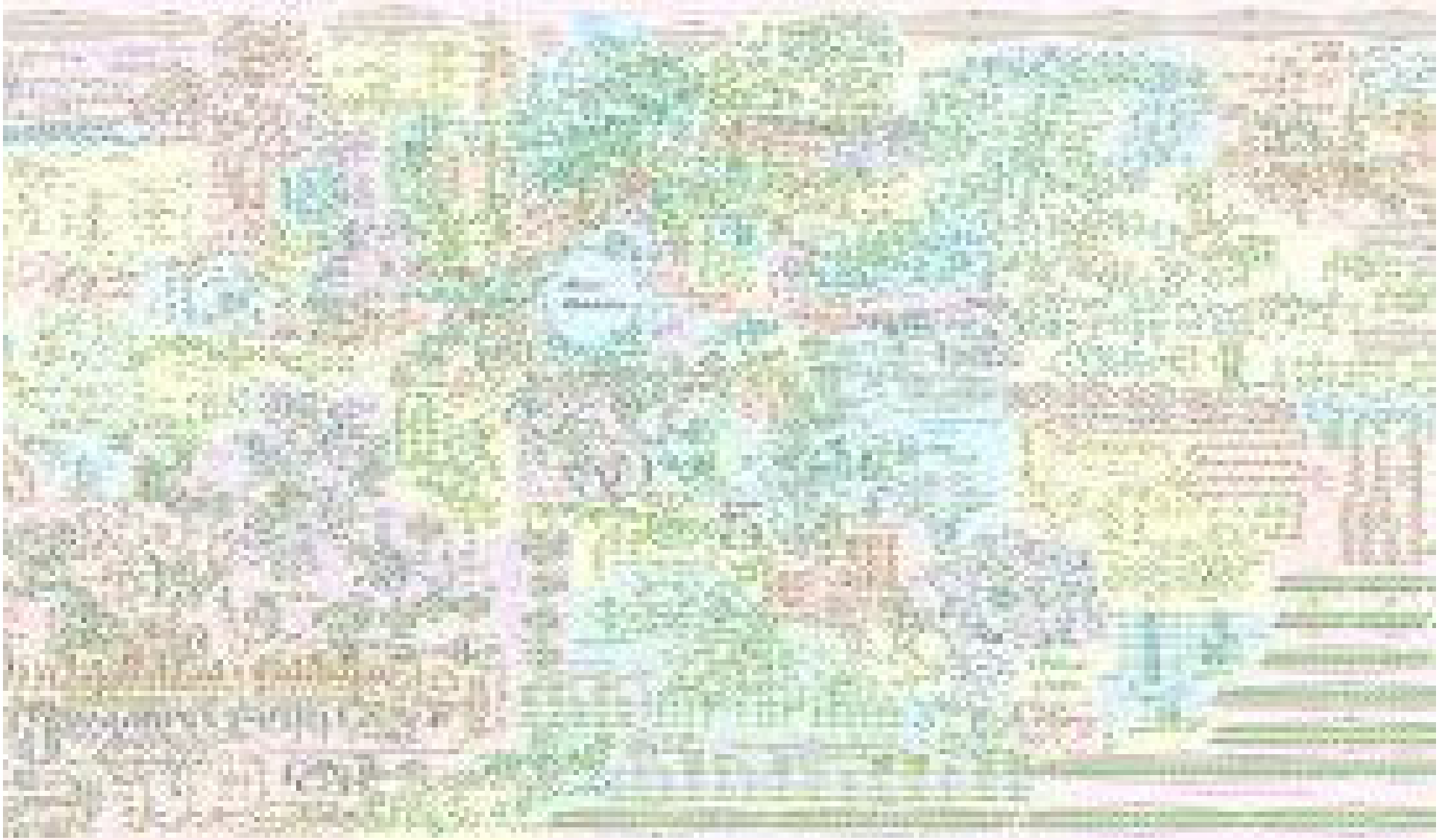
BiGCaT

department of bioinformatics

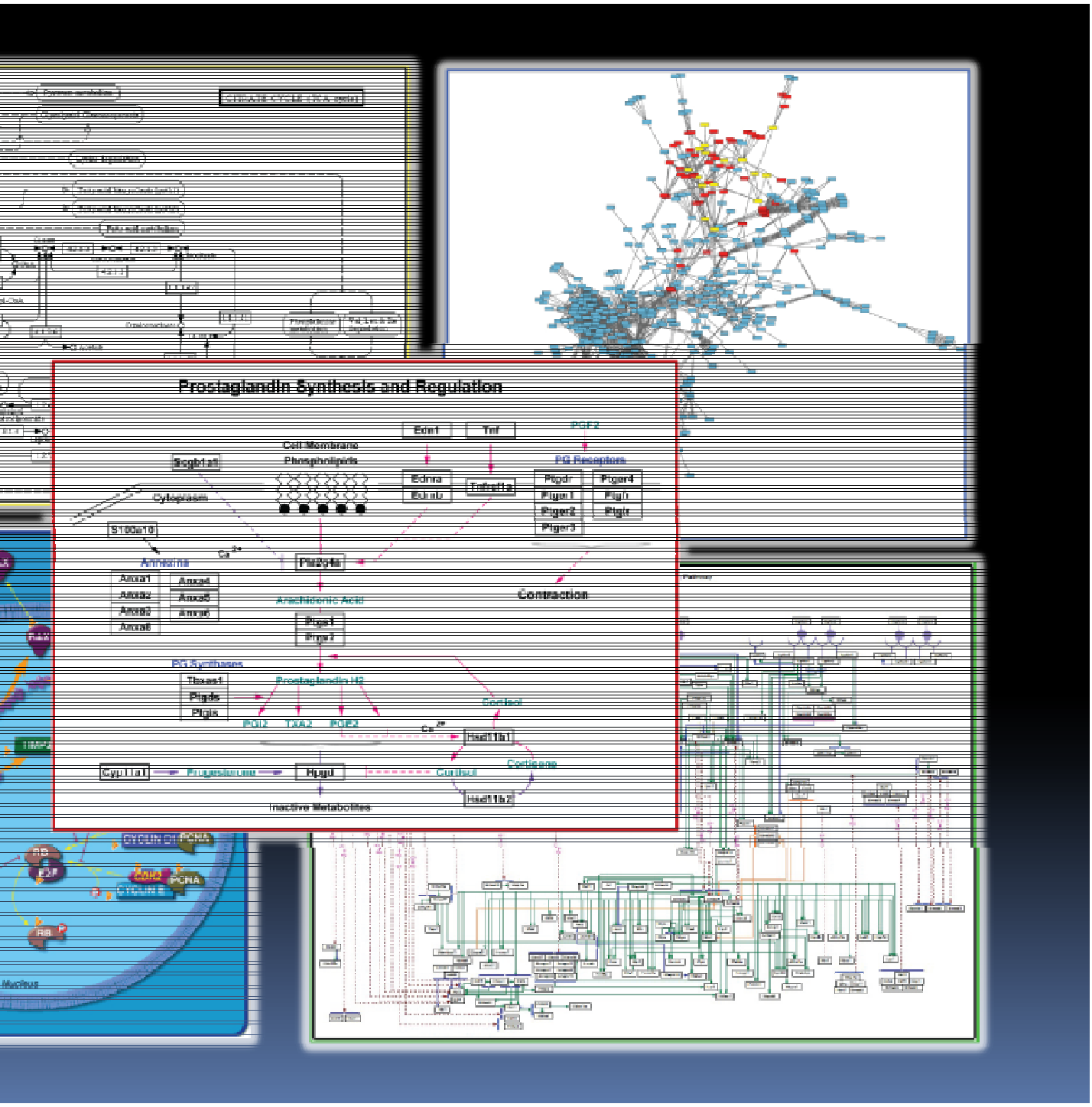
Maastricht University

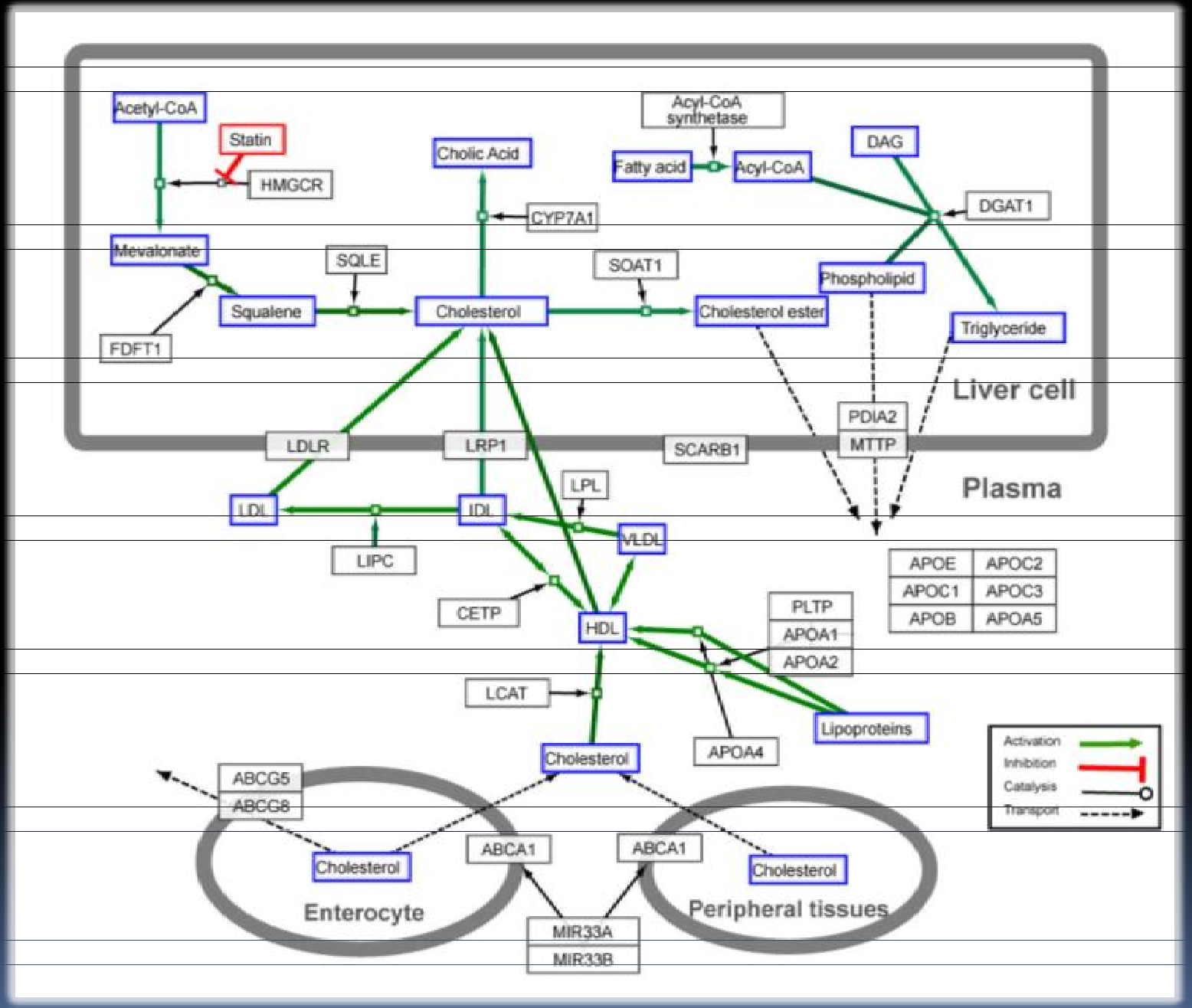

Gerhard Michal 1974

Recon2 a Google map of human metabolism

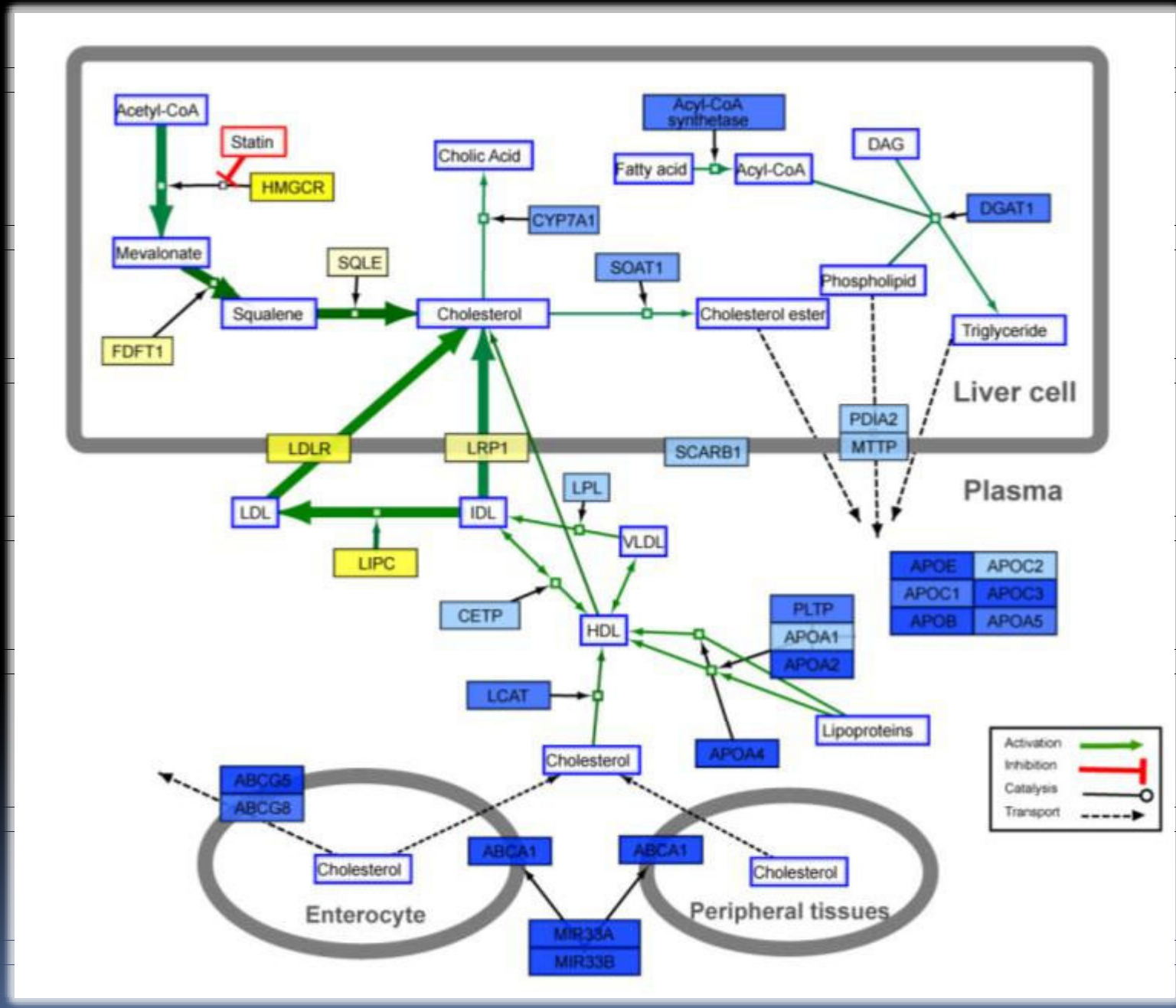


Collaborating Systems Biology and Metabolomics groups: 2013

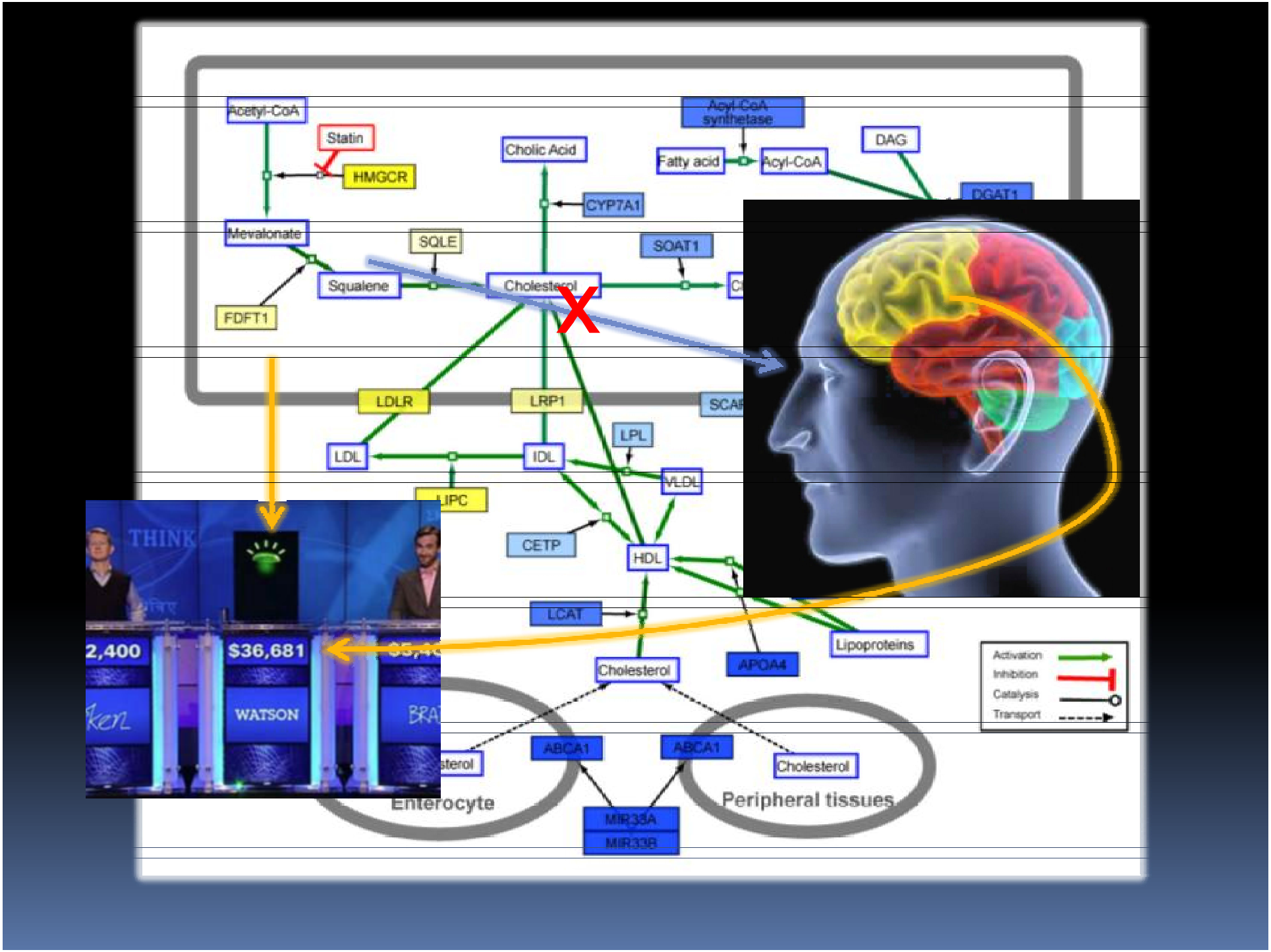

Prostaglandin Synthesis and Regulation
Edn1
Tnf
PGE2
Cell Membrane
Phospholipids
PG Receptors
Cytoplasm
S100a10
Ca2+
Annexins
Anxa1
Anxa2
Anxa3
Anxa4
Anxa5
Anxa6
Arachidonic Acid
Ptgs1
Ptgs2
PG Synthases
Tbxas1
Ptgds
Ptgis
Prostaglandin H2
PGD2
TXA2
PGF2
Contract
Hsd11b1
Cortisone
Hsd11b2
Cyp11a1
Progesterone
Hpgd
Inactive Metabolites
Contraction
CYCLIN D1
CDK4
PCNA
CYCLIN E
RB
E2F
Nucleus

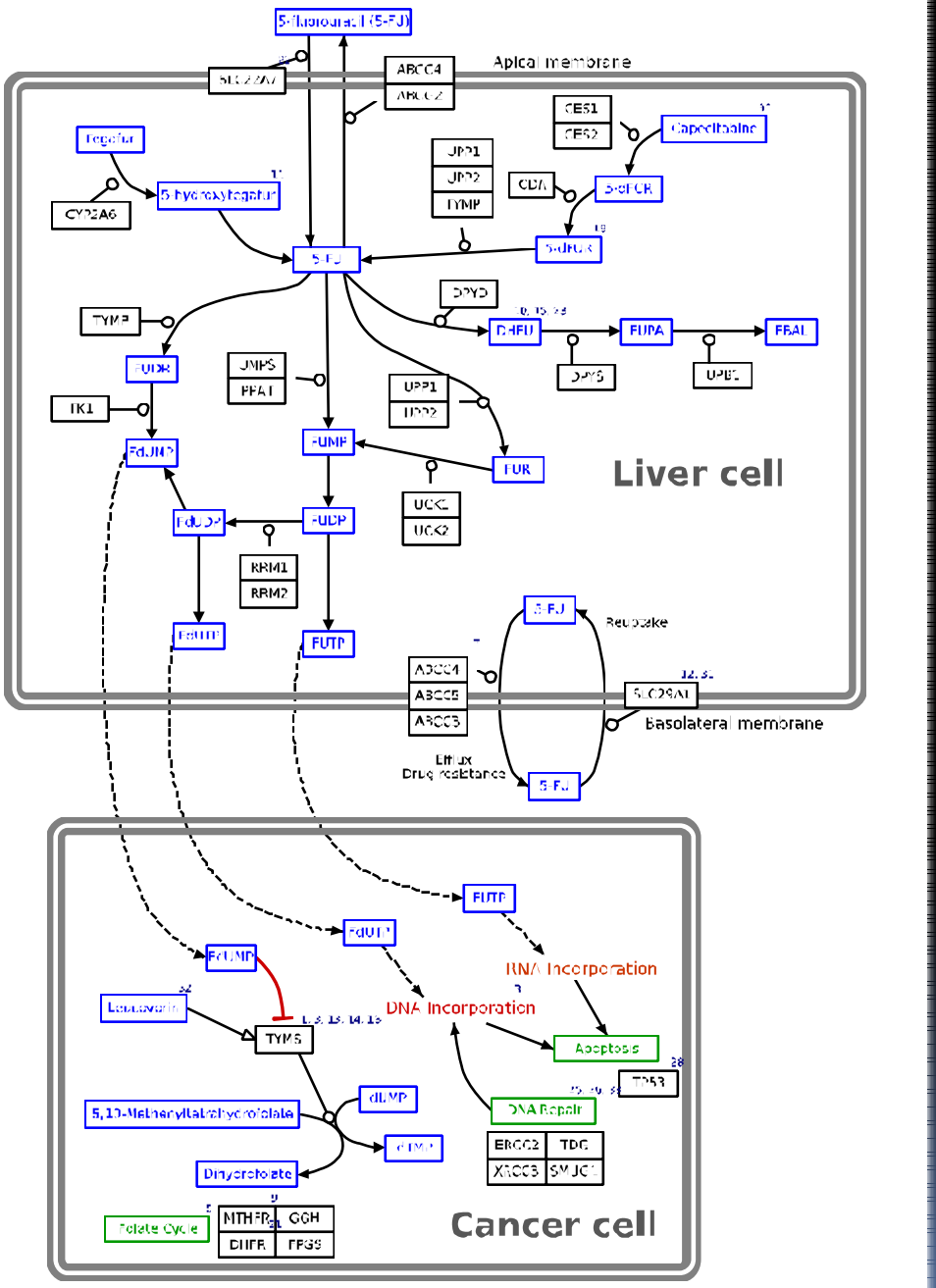


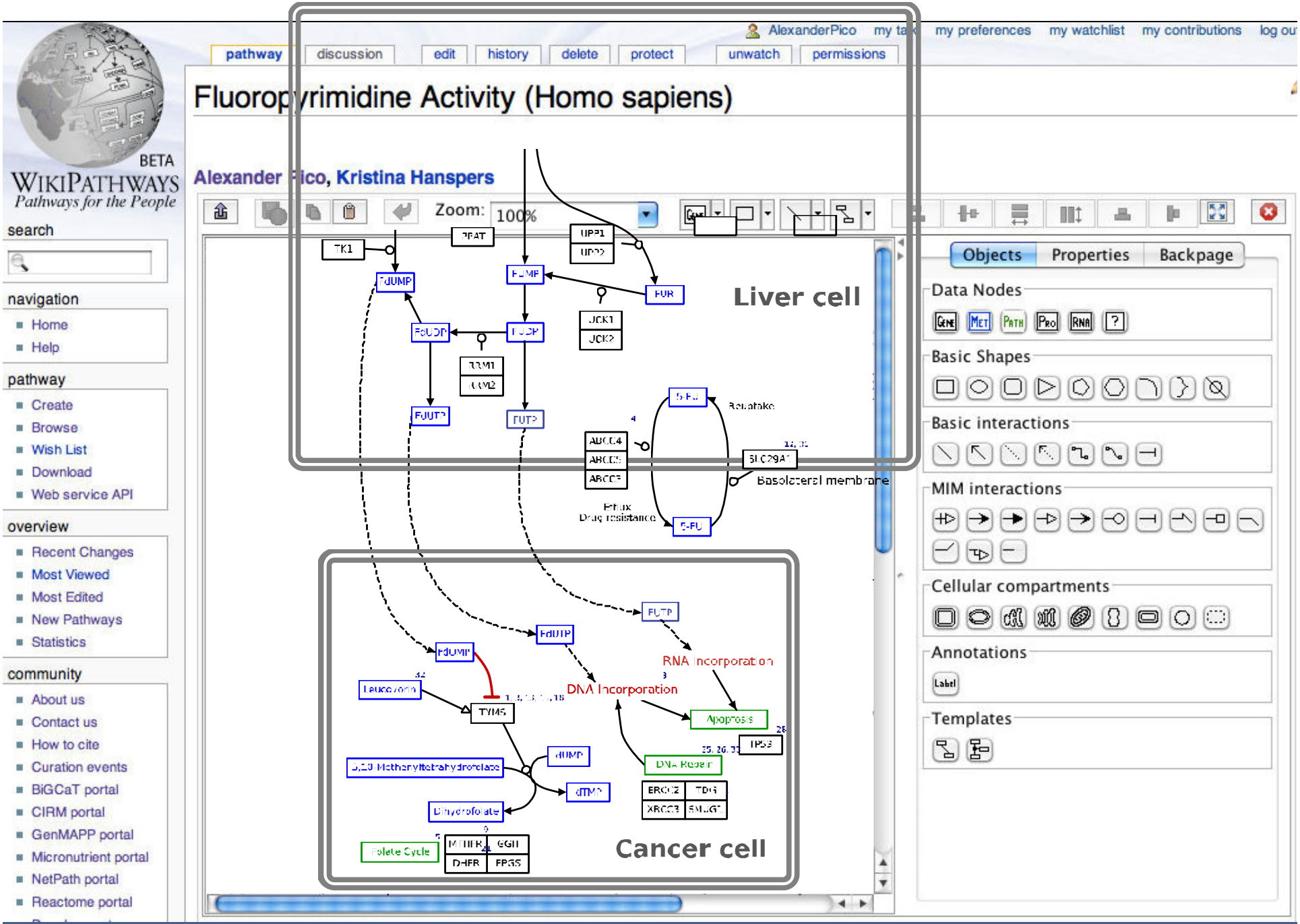
http://www.wikipathways.org/index.php/Pathway:WP430



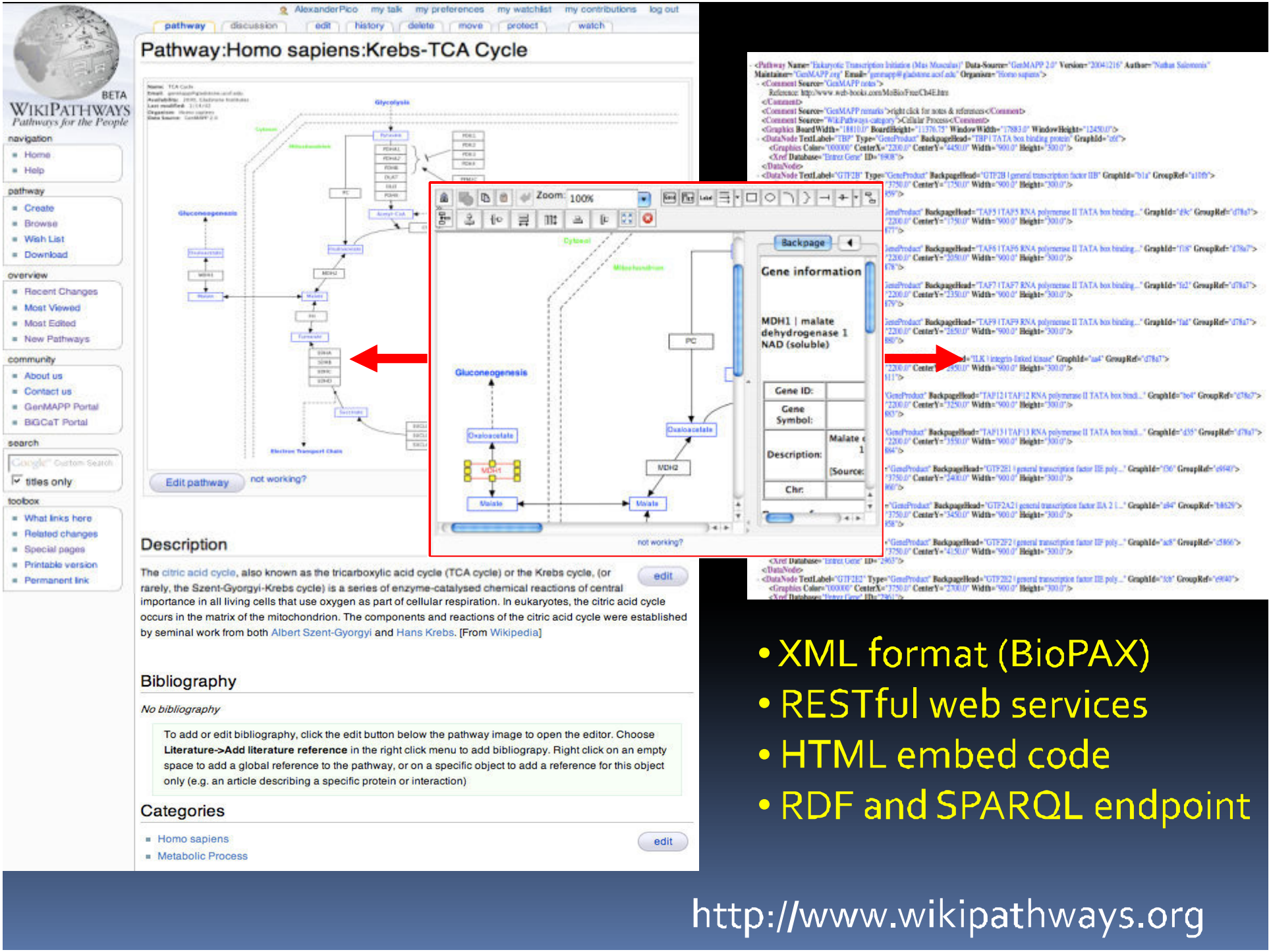
http://www.wikipathways.org/index.php/Pathway:WP430

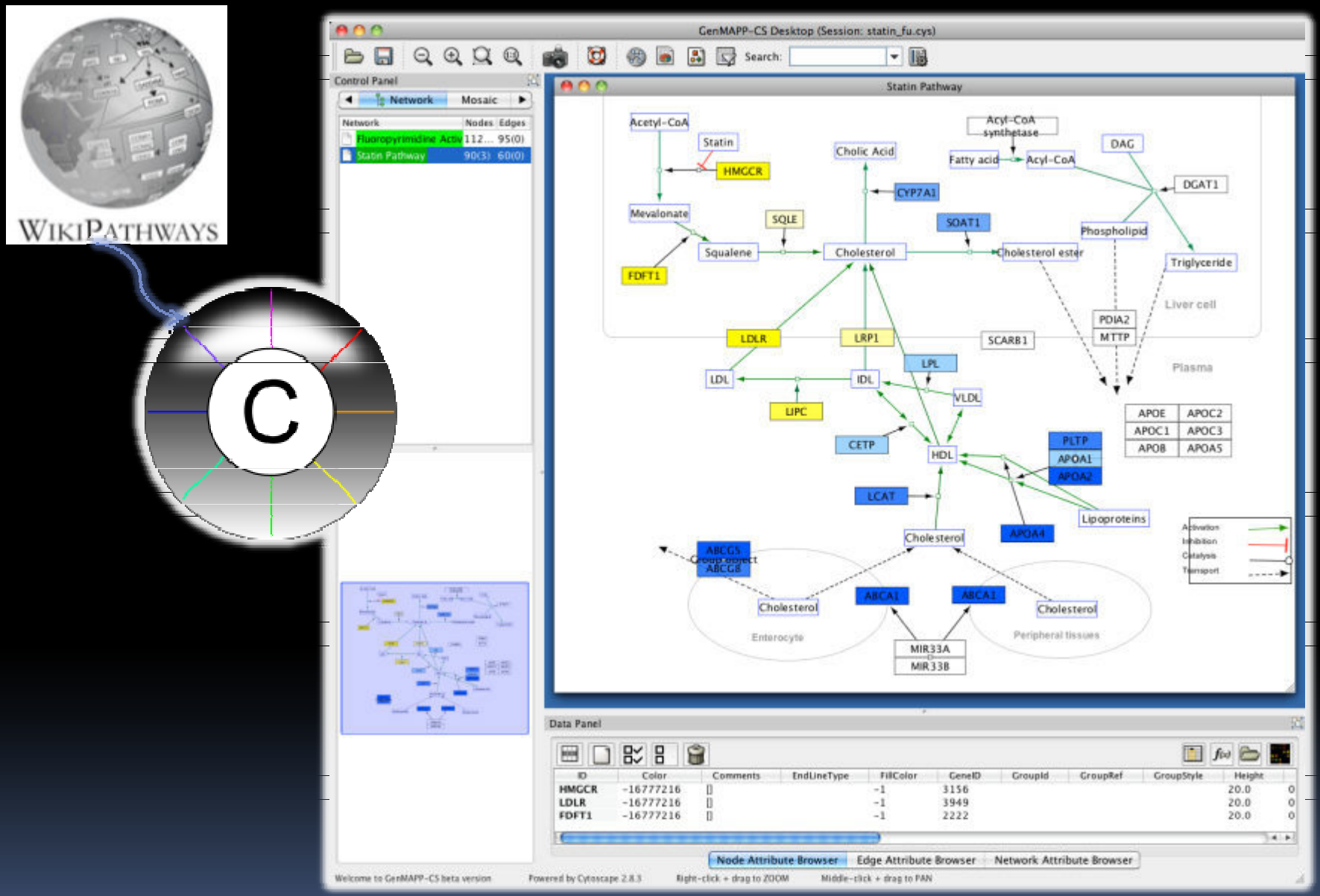

Acetyl-CoA
Statin
HMGCR
Mevalonate
FDFT1
Squalene
SQLE
Cholic Acid
CYP7A1
Cholesterol
Acyl-CoA synthetase
Fatty acid
Acyl-CoA
DAG
SOAT1
LDLR
LRP1
LDL
IDL
LPL
VLDL
LIPC
CETP
HDL
LCAT
Lipoproteins
APOA4
Cholesterol
ABCA1
ABCA1
Cholesterol
MIR33A
MIR33B
Enterocyte
Peripheral tissues
Activation
Inhibition
Catalysis
Transport
THINK
2,400
$36,681
WATSON


5-fluorouracil (5-FU)
Apical membrane
ABCC4
ABCC2
CES1
CES2
Capecitabine
Tegafur
5-hydroxytegafur
CYP2A6
UP1
UP2
TYMP
CDA
5-dFCR
5-dFUR
5-FU
DPYD
TYMP
DHFU
FUPA
FBAL
DPYS
UPB1
UPDR
UMPS
PPAT
UPP1
UPP2
UR1
FdUMP
FUMP
FUR
Liver cell
FdUDP
FUDP
UCKL
UCK2
RRM1
RRM2
FdUTP
FUTP
5-FU
Reuptake
ABCC4
ABCC5
ABCC3
SLCO2A1
Basolateral membrane
Efflux
Drug resistance
5-FU
FUTP
FdUTP
FdUMP
RNA Incorporation
DNA Incorporation
Leucovorin
TYMS
Apoptosis
TP53
5,10-Methenyltetrahydrofolate
dUMP
DNA Repair
ERCC2
TDG
XRCC3
SMUC1
dTMP
Dihydrofolate
Folate Cycle
MTHFR
GGH
DHFR
FPGS
Cancer cell

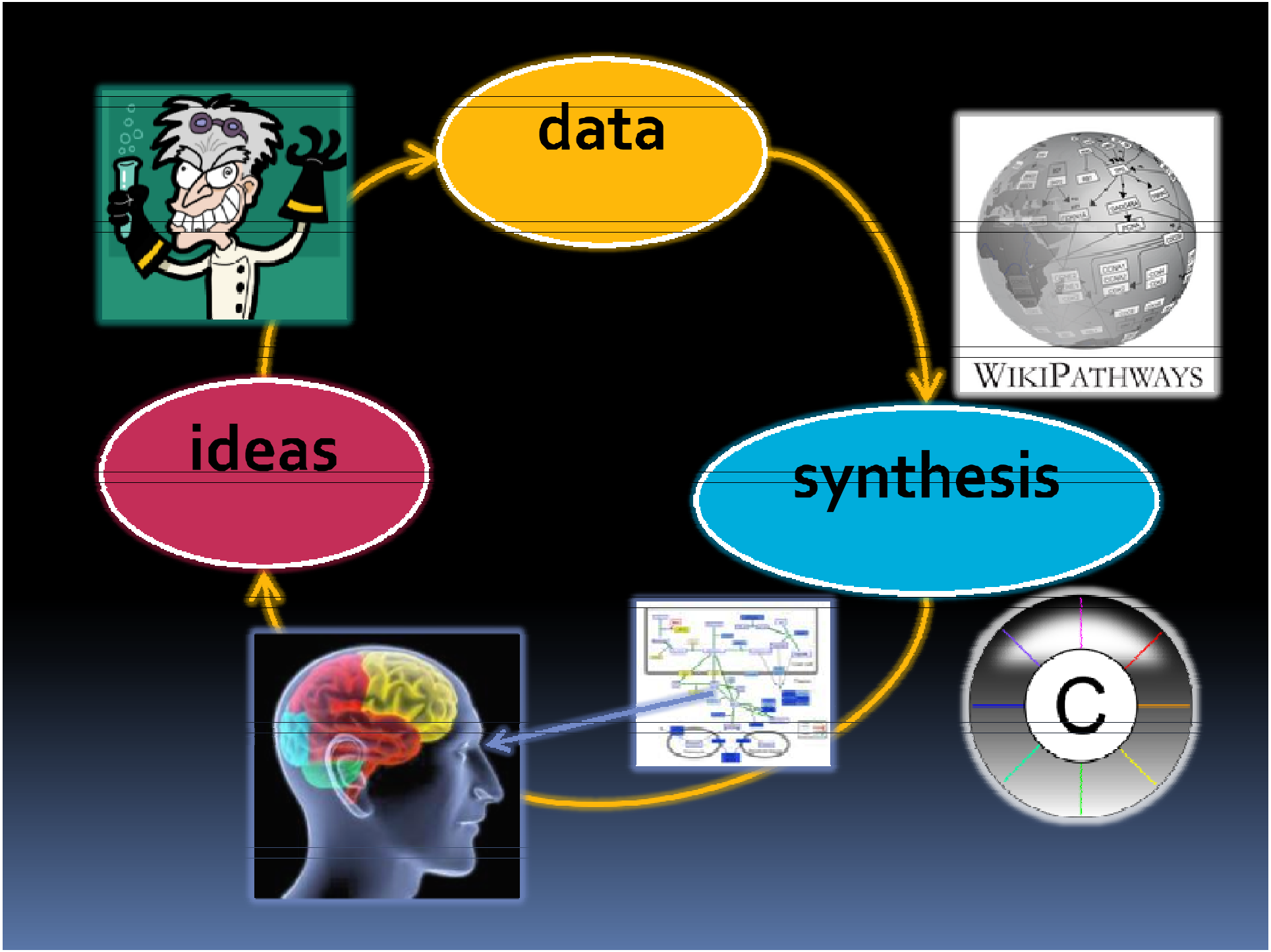


http://www.wikipathways.org/index.php/Pathway:WP1601


BETA
WIKIPATHWAYS
Pathways for the People
navigation
Home
Help
pathway
Create
Browse
Wish List
Download
overview
Recent Changes
Most Viewed
Most Edited
New Pathways
community
About us
Contact us
GenMAPP Portal
BiGCaT Portal
search
titles only
toolbox
What links here
Related changes
Special pages
Printable version
Permanent link
AlexanderPico my talk my preferences my watchlist my contributions log out
pathway discussion edit history delete move protect watch
Pathway:Homo sapiens:Krebs-TCA Cycle
Edit pathway not working?
Description
The citric acid cycle, also known as the tricarboxylic acid cycle (TCA cycle) or the Krebs cycle, (or rarely, the Szent-Gyorgyi-Krebs cycle) is a series of enzyme-catalysed chemical reactions of central importance in all living cells that use oxygen as part of cellular respiration. In eukaryotes, the citric acid cycle occurs in the matrix of the mitochondrion. The components and reactions of the citric acid cycle were established by seminal work from both Albert Szent-Gyorgyi and Hans Krebs. [From Wikipedia]
edit
Bibliography
No bibliography
To add or edit bibliography, click the edit button below the pathway image to open the editor. Choose Literature->Add literature reference in the right click menu to add bibliograpy. Right click on an empty space to add a global reference to the pathway, or on a specific object to add a reference for this object only (e.g. an article describing a specific protein or interaction)
Categories
Homo sapiens
Metabolic Process
edit
Zoom: 100%
Backpage
Gene information
MDH1 | malate dehydrogenase 1 NAD (soluble)
Gene ID:
Gene Symbol:
Description:
Chr:
Gluconeogenesis
Oxaloacetate
MDH1
Malate
PC
MDH2
not working?
- Pathway Name="Eukaryotic Transcription Initiation (Mus Musculus)" Data-Source="GenMAPP 2.0" Version="20041216" Author="Nathan Salomonis" Maintainer="GenMAPP.org" Email="genmapp@gladstone.ucsf.edu" Organism="Homo sapiens"
- Comment Source="GenMAPP notes"
Reference: http://www.web-books.com/MoBio/FreeCh4E.htm
/Comment
Comment Source="GenMAPP remarks">right click for notes & references</Comment
Comment Source="WikiPathways-category">Cellular Process</Comment
Graphics BoardWidth="18810.0" BoardHeight="11376.75" WindowWidth="17983.0" WindowHeight="12450.0"/
- DataNode TextLabel="TBP" Type="GeneProduct" BackpageHead="TBP | TATA box binding protein" GraphId="c6f"
Graphics Color="000000" CenterX="2200.0" CenterY="4450.0" Width="900.0" Height="300.0"/
Xref Database="Entrez Gene" ID="6908"/
/DataNode
- DataNode TextLabel="GTF2B" Type="GeneProduct" BackpageHead="GTF2B | general transcription factor IIB" GraphId="b1a" GroupRef="a10f0"
"3750.0" CenterY="1750.0" Width="900.0" Height="300.0"/
856"/
- DataNode TextLabel="GTF2E1" Type="GeneProduct" BackpageHead="GTF2E1 | general transcription factor IIE poly..." GraphId="f50" GroupRef="d9f40"
"3750.0" CenterY="2400.0" Width="900.0" Height="300.0"/
860"/
- DataNode TextLabel="GTF2A2" Type="GeneProduct" BackpageHead="GTF2A2 | general transcription factor IIA 2 (..." GraphId="a94" GroupRef="b8529"
"3750.0" CenterY="3450.0" Width="900.0" Height="300.0"/
858"/
- DataNode TextLabel="GTF2F2" Type="GeneProduct" BackpageHead="GTF2F2 | general transcription factor IIF poly..." GraphId="ac8" GroupRef="c5866"
"3750.0" CenterY="4150.0" Width="900.0" Height="300.0"/
Xref Database="Entrez Gene" ID="2963"/
/DataNode
- DataNode TextLabel="GTF2E2" Type="GeneProduct" BackpageHead="GTF2E2 | general transcription factor IIE poly..." GraphId="fcb" GroupRef="d9f40"
Graphics Color="000000" CenterX="3750.0" CenterY="2700.0" Width="900.0" Height="300.0"/
Xref Database="Entrez Gene" ID="2961"/
• XML format (BioPAX)
• RESTful web services
• HTML embed code
• RDF and SPARQL endpoint
http://www.wikipathways.org

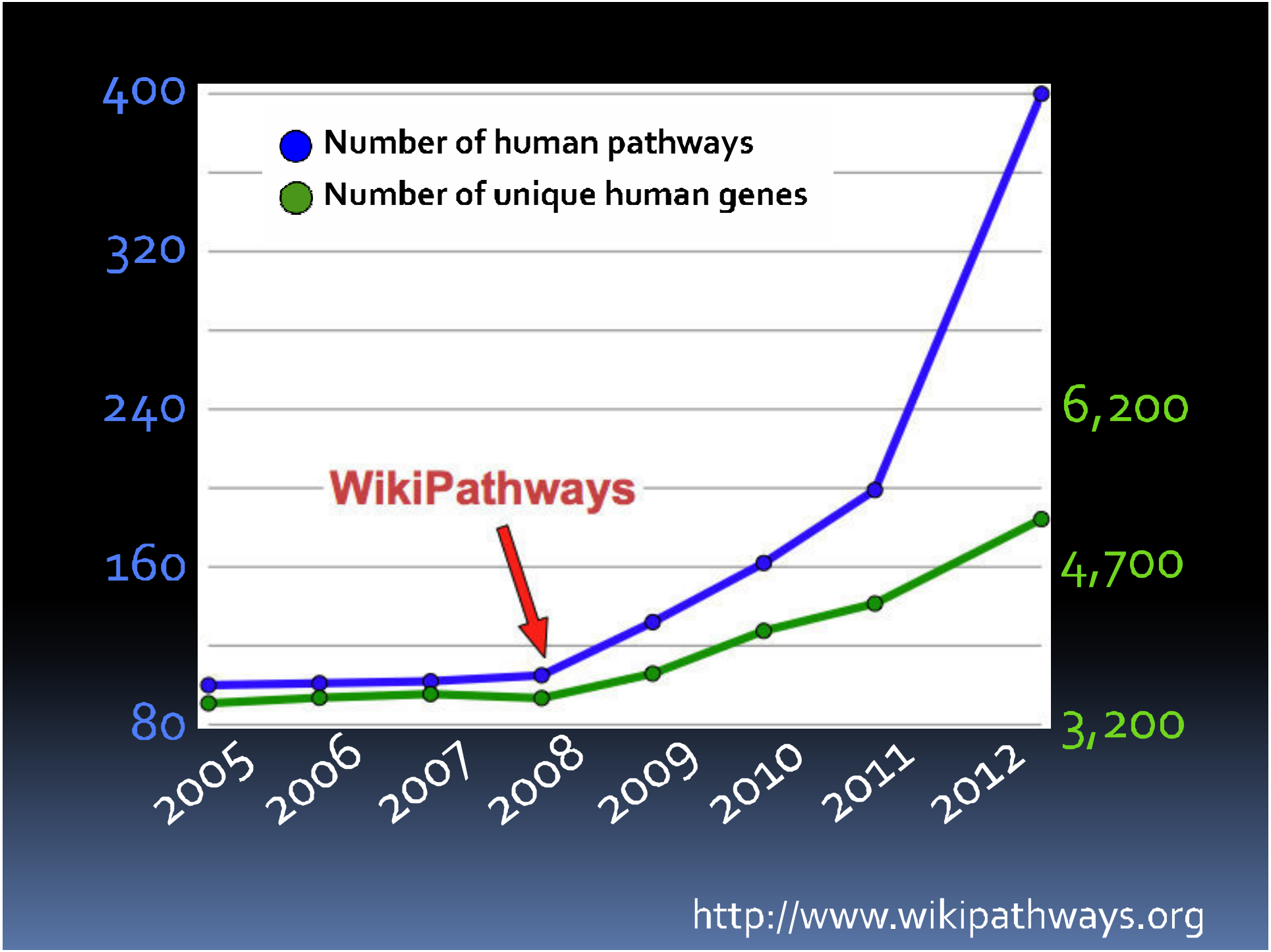


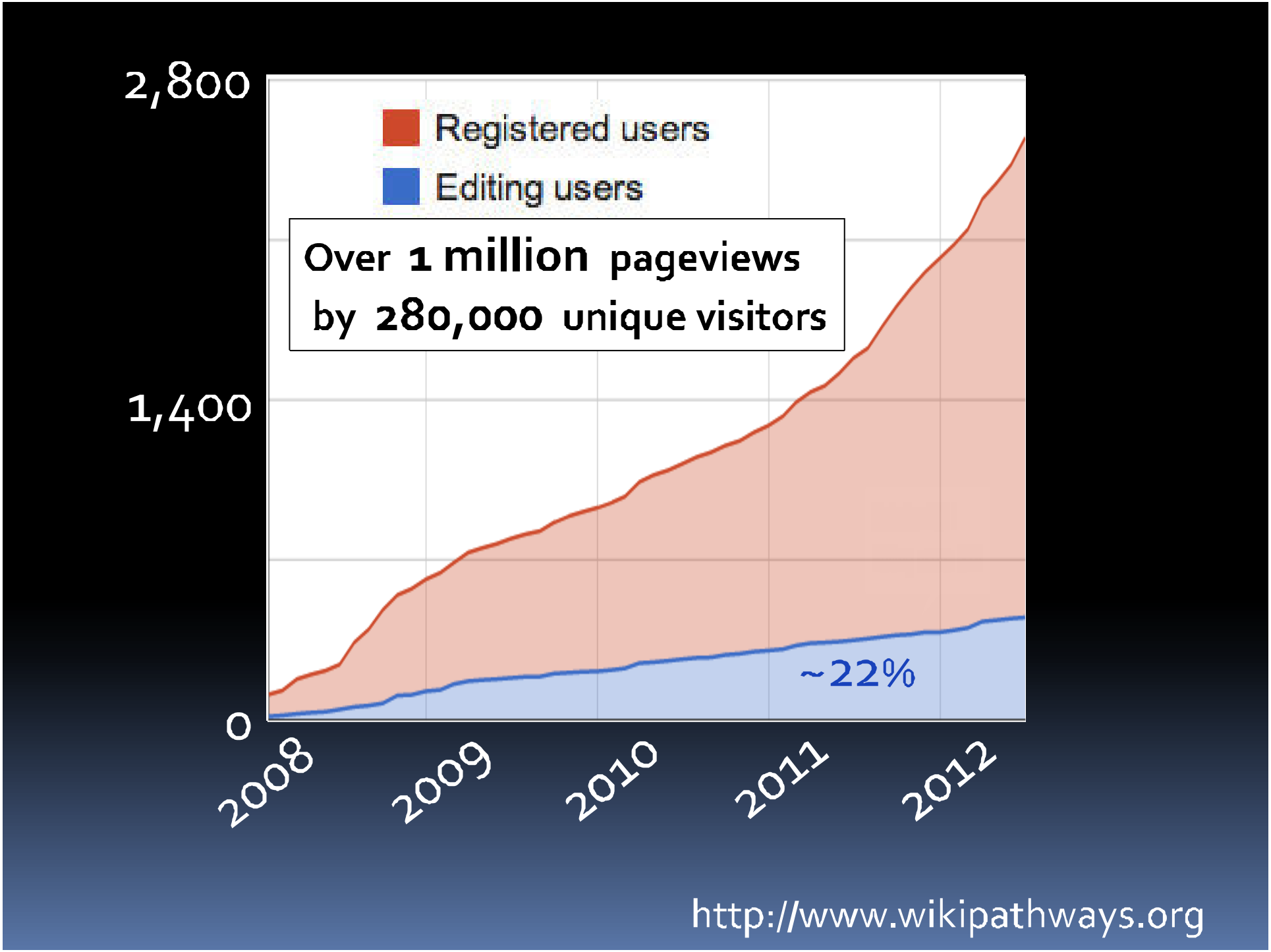
http://www.cytoscape.org


data
WIKIPATHWAYS
ideas
synthesis
C

Set Early Milestones

- Online (Mar '07) ✓*Success!*
- First unknown user (Jan '08)


400
320
240
160
80
Number of human pathways
Number of unique human genes
WikiPathways
6,200
4,700
3,200
2005
2006
2007
2008
2009
2010
2011
2012
http://www.wikipathways.org


2,800
1,400
0
Registered users
Editing users
Over 1 million pageviews
by 280,000 unique visitors
~22%
2008
2009
2010
2011
2012
http://www.wikipathways.org

Don’t Try to Change the World

Work with (not against) established:

- Models
- Communities
- Tools and pipelines
- Publishing models

Go Ahead, Change the World

- Tweak established models
- Grow communities
- Change perspectives
  - *everyone is a curator*
  - *knowledge should be open*

Go Ahead, Change the World

- Tweak established models
- Grow communities
- Change perspectives
- New attribution systems
  - *redefine "publication"*
  - *redefine "productive"*

Go Ahead, Change the World

- Tweak established models
- Grow communities
- Change perspectives
- New attribution systems
- New analysis pipelines
  - *connect with other community-curated resources*

Search

nd analyze biological pathways

t/Help Downloads Plugins Cite Us About

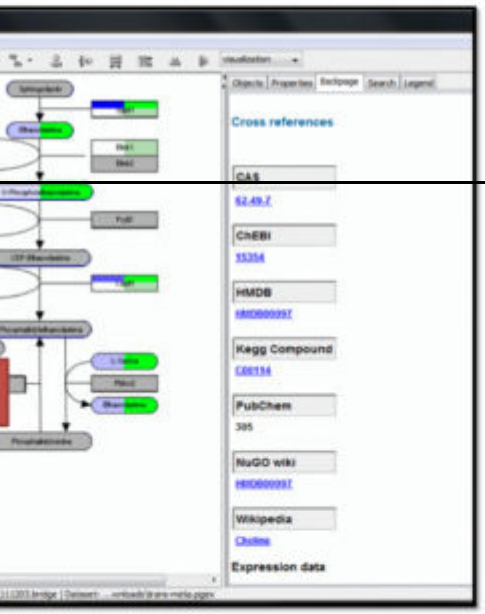
l pathway analysis
nd analyze

How to use PathVisio?

Learn how you can download PathVisio and go through the tutorials to find out how to analyze and visualize your data. Get started.

PathVisio Plugins

Plugins are extensions that provide advanced analysis methods, visualization options or additional import/export functionality. Find out more.



News

- **PathVisio @ OpenTox Euro 2013:** Martina Kutmon is giving a presentation on PathVisio 3 and how PathVisio is used in the field of Toxicology.
- **GSoC 2013:** All three students finished their GSoC 2013 project successfully! Congratulations! Check out their blogs here.
- **Developer website:** The previous website will stay available on developers.pathvisio.org and will be used as a developers website.
- **Plugin repository:** Find plugins in the plugin repository and install them through the plugin manager.

Releases

- **Oct 29 2013 PathVisio 3.1.2 has been released!**
- **Oct 04 2013 PathVisio 3.1.1 has been released!**
- **Jul 02 2013 PathVisio 3.1.0 has been released!**
- **Jun 06 2013 PathVisio 3.0.1 has been released!**

Check out the detailed release notes on our developers website.

Professional Open Source

- Subversion source repository
- License!
- Development web site
- Bug tracker
- Mailing lists
- Development and Release plans
- Modular (plugins, OSGi)

The New Agilent

CORE TECHNOLOGY PLATFORMS

Gas chromatography | Liquid Chromatography | Mass spectrometry | Spectroscopy | NMR Spectroscopy | Automation | Software | Chemistries | Immunohistochemistry | FISH probes | Microarrays | Target enrichment | Bioreagents | Services

Key Academic Innovation Partners

About 100 active university collaborations annually

Agilent Thought Leader Program

http://www.agilent.com/univ_relation/TLP/index.shtml

Thought Leader Awards

Promote fundamental advances in life science through contribution to research of thought leaders

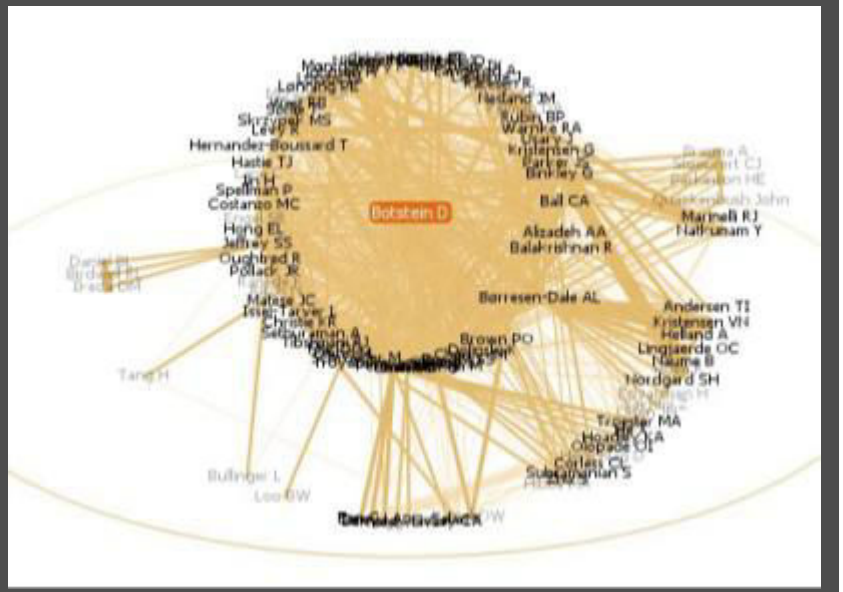
- Align societal trends, academic research and rapidly advancing Agilent measurement platforms:

  Synthetic Biology, Structural Biology, OMICS & Integrated Biology, *in vitro* Toxicology, and Environemntal & Food Safety

- Candidates selected based on scientific leadership, productivity, project significance (invitational program)

- High-level executive sponsorship and active support throughout Agilent enable breakthrough research

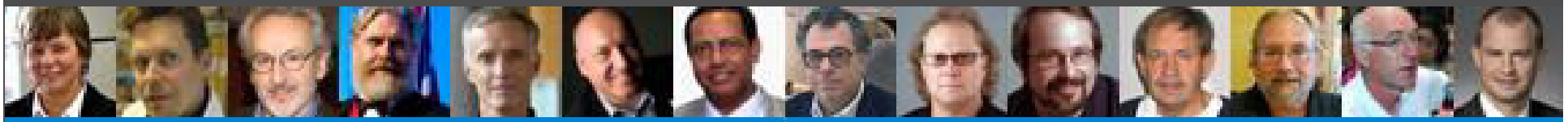
TL Network



Early Career Professor Award

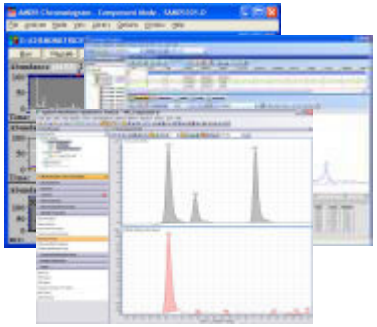
Establish strong collaborative relationships with highly promising early career professors

2013 focus:
Contributions to cancer diagnostics



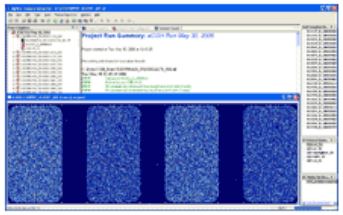
Agilent Integrated Biology Workflows

LC/MS
GC/MS



MassHunter Qual/Quant
ChemStation AMDIS

Microarrays

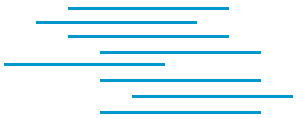


Feature Extraction

GeneSpring Platform

Biological
Pathways

NGS



Alignment to Reference Genome

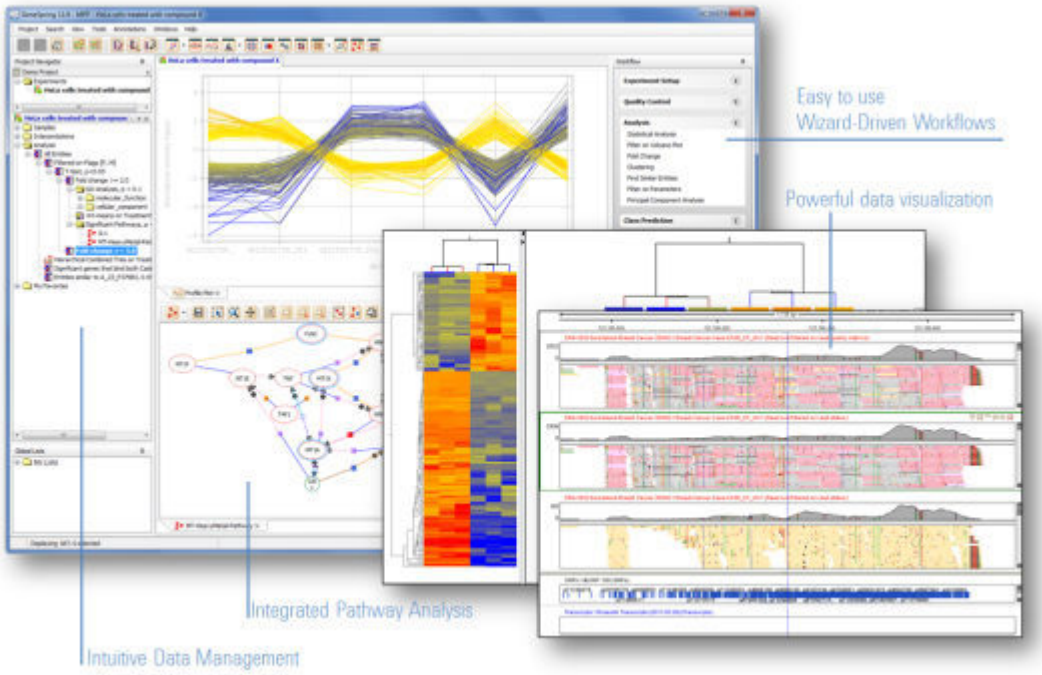
GeneSpring

Agilent's Platform for Multi-Omics Data Analysis

10 More than Years of Experience 100 Hundr 10000 References In Google Scholar

Agilent Technologies

GeneSpring modules



GX (Gene Expression)

mRNA, miRNA, Exon arrays
GWAS, CNV via SNP arrays

NGS (Next-Gen Sequencing)

SureSelect Target Enrichment
Whole Genome Sequencing

MPP (Mass Profiler Pro)

Proteomics
Metabolomics

PA (Pathway Analysis)

GeneSpring modules

GX

mRNA
microRNA
QPCR
Alternative Splicing
GWAS & CNV via SNP arrays

GX (Gene Expression)

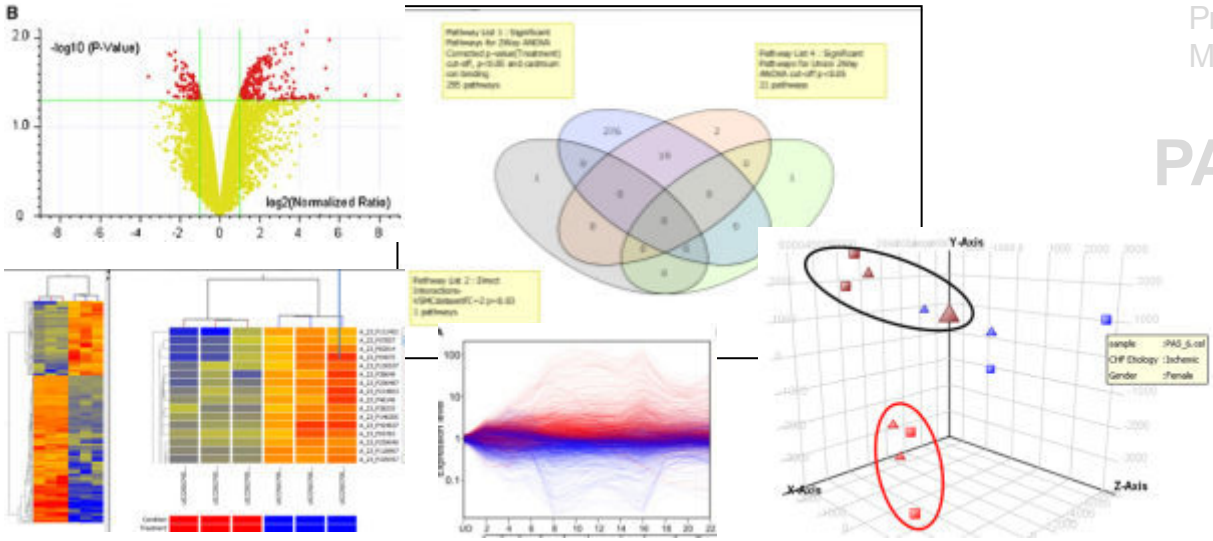
NGS (Next-Gen Sequencing)

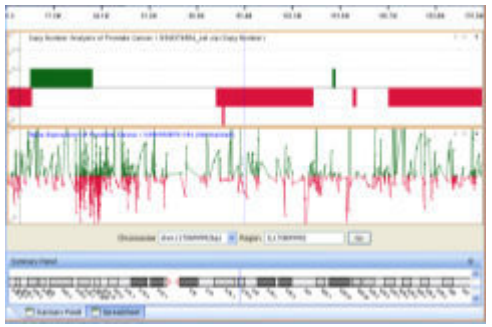
SureSelect Target Enrichment
Whole Genome Sequencing

MPP (Mass Profiler Pro)

Proteomics
Metabolomics

PA (Pathway Analysis)





GeneSpring modules

NGS

SureSelect Target Enrichment
Whole Genome Sequencing
DNA-Seq
RNA-Seq
Methyl-Seq
ChIP-Seq
small RNA-Seq

GX (Gene Expression)

NGS (Next-Gen Sequencing)
SureSelect Target Enrichment
Whole Genome Sequencing

MPP (Mass Profiler Pro)
Proteomics
Metabolomics

PA (Pathway Analysis)

NGS Analysis Workflow:

1. Align data
2. Load BAM/SAM into GS NGS
3. Measure Gene Expression, Find variants, methyl calls
4. Biological Contextualization (Integrated Genomics, GO, Pathways)

GeneSpring modules

MPP

- Import, store, and visualize Agilent Metabolomics & Proteomics data (LC/MS, GC/MS)
- Generic file import
- Statistical analysis
- ID Browser for compound identification

GX (Gene Expression)

NGS (Next-Gen Sequencing)

SureSelect Target Enrichment
Whole Genome Sequencing

MPP (Mass Profiler Pro)

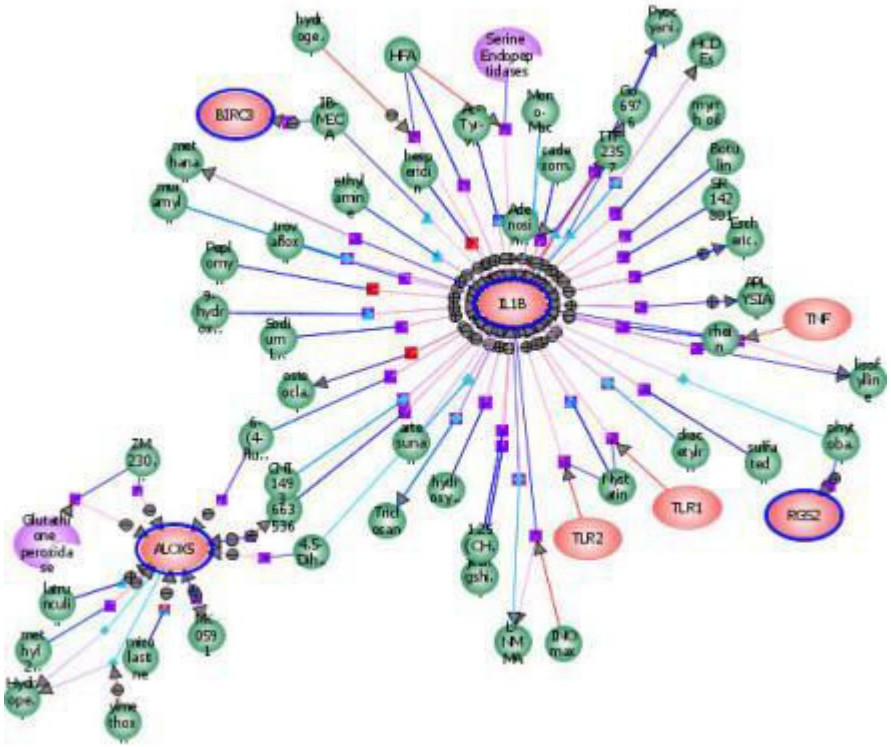
Proteomics
Metabolomics

PA (Pathway Analysis)

GeneSpring modules

PA

Multi-Omic Analysis
Canonical Pathways
Network Discovery



GX (Gene Expression)

NGS (Next-Gen Sequencing)

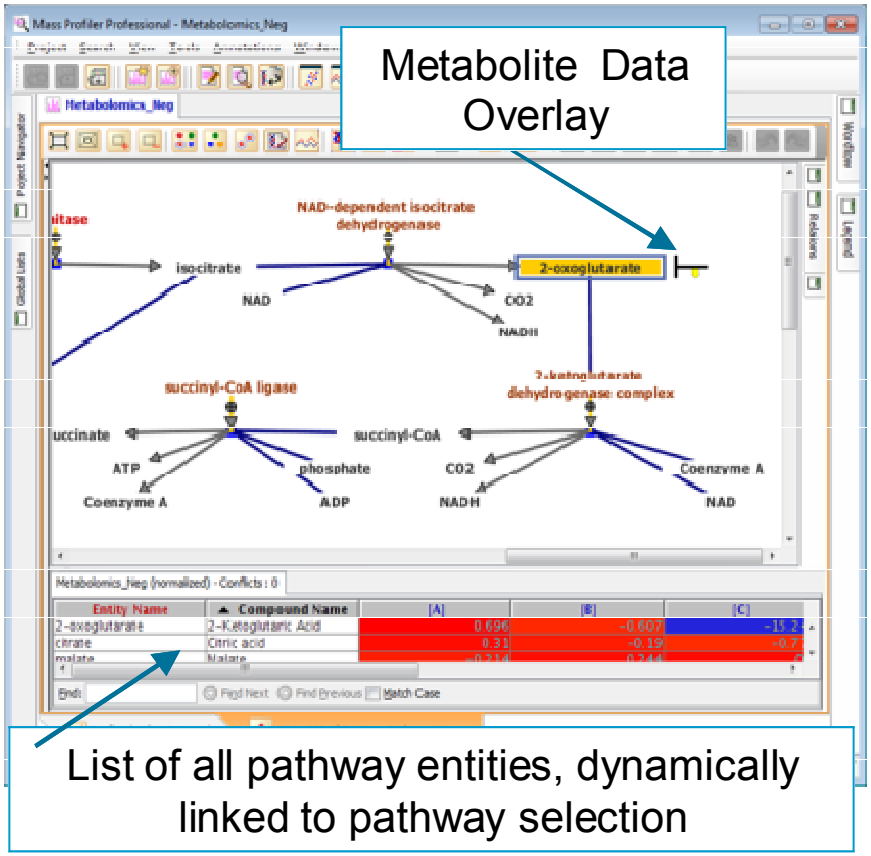
SureSelect Target Enrichment
Whole Genome Sequencing

MPP (Mass Profiler Pro)

Proteomics
Metabolomics

PA (Pathway Analysis)

Pathway Architect



- Map and visualize data from one or two types of –omic data on pathways
- Search, browse and filter pathways

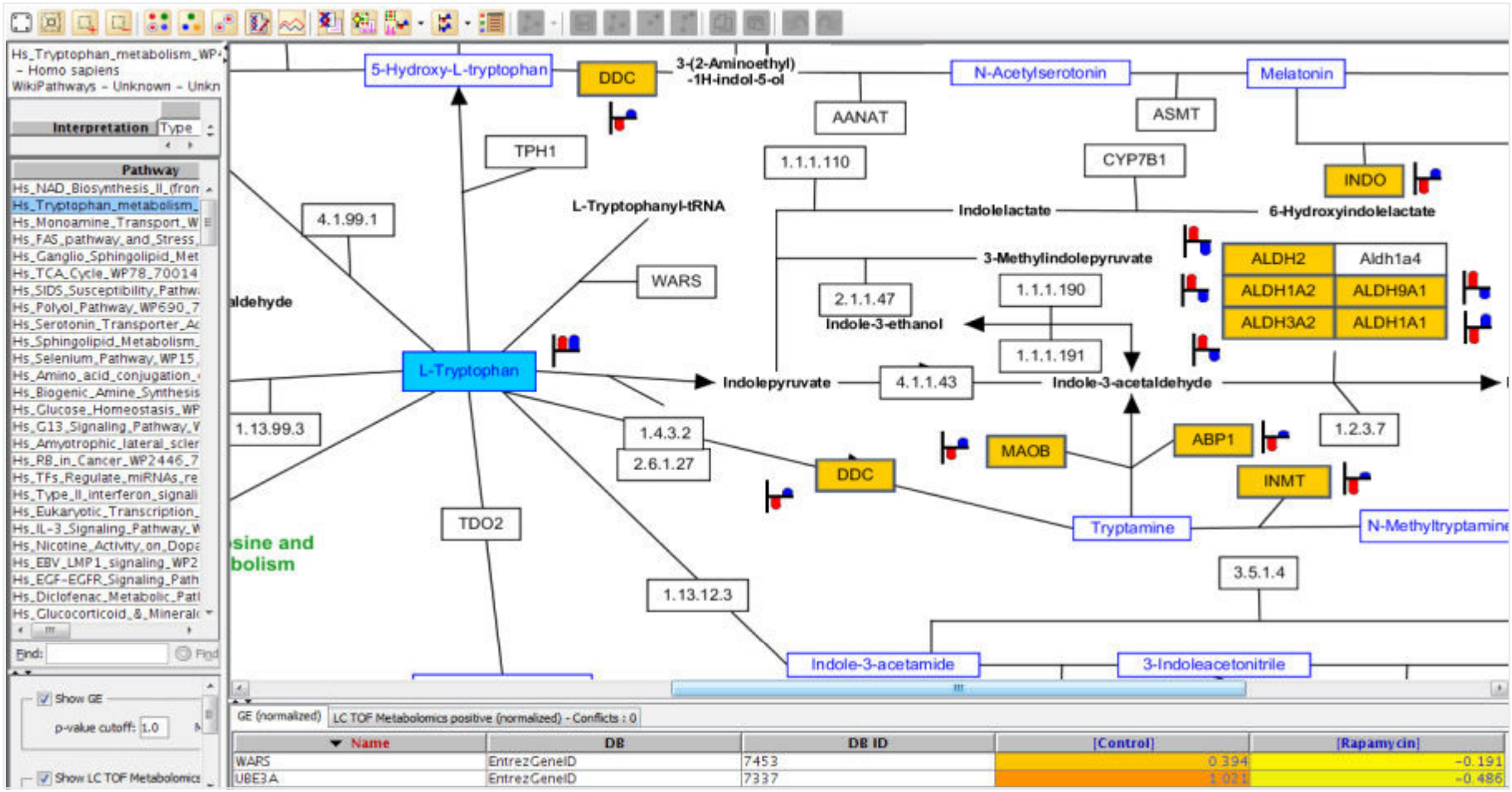
Supports pathways from:

- WikiPathways
- BioCyc

- Supported pathway formats
- BioPAX 3 – Pathway Commons, Reactome, NCI Nature Pathway
- GPML – PathVisio –custom drawing

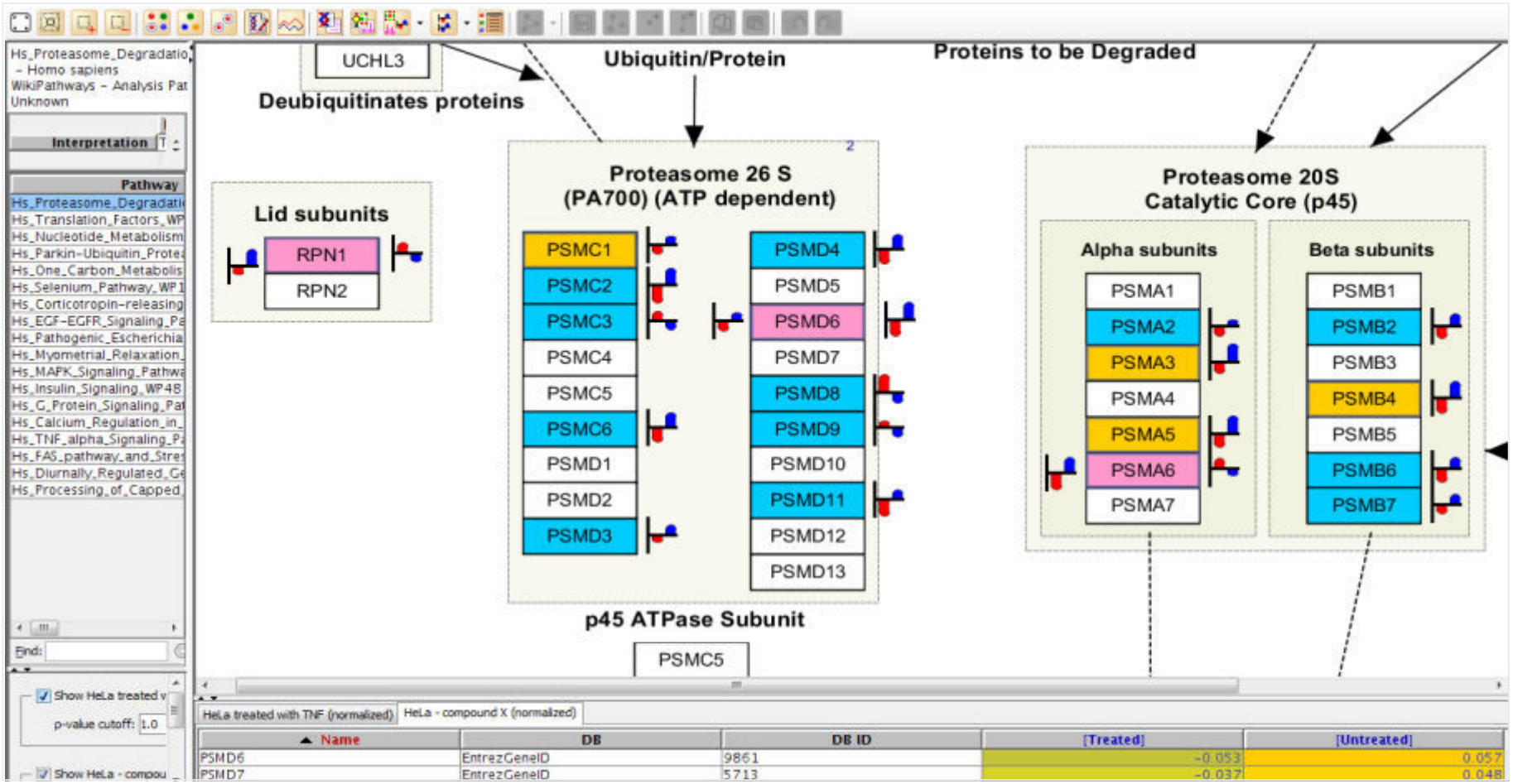
Export compound list from pathways

Genomics and Metabolomics data overlay on Tryptophan Metabolism WikiPathway



metabolites proteins

Genomics and Proteomics data overlay on Proteasome Degradation WikiPathway



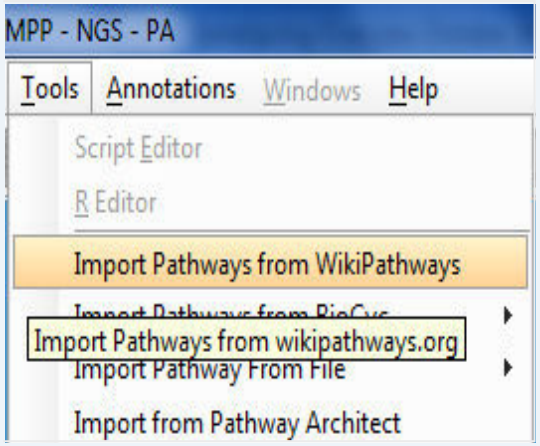
Proteomics | Genomics | Proteomics and Genomics

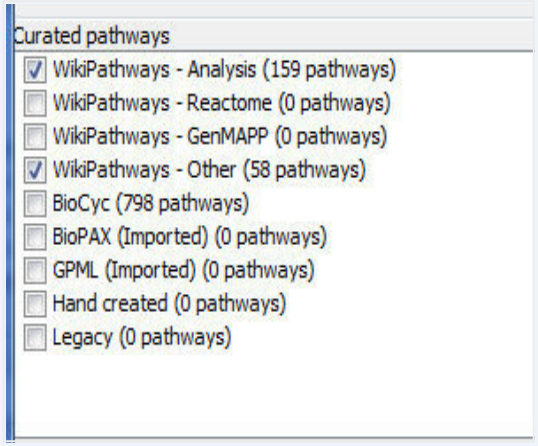
Analysis with WikiPathways in Pathway Architect

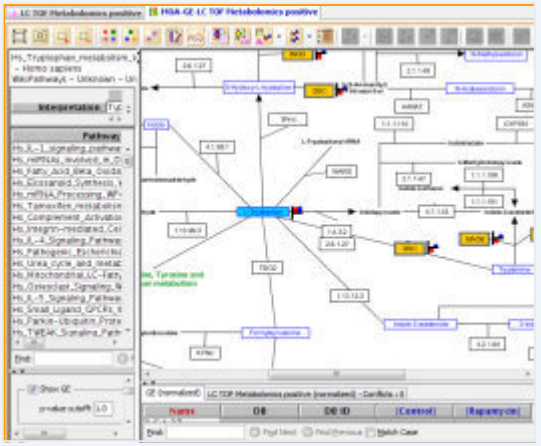
Import of WikiPathways

Select WikiPathways for analysis

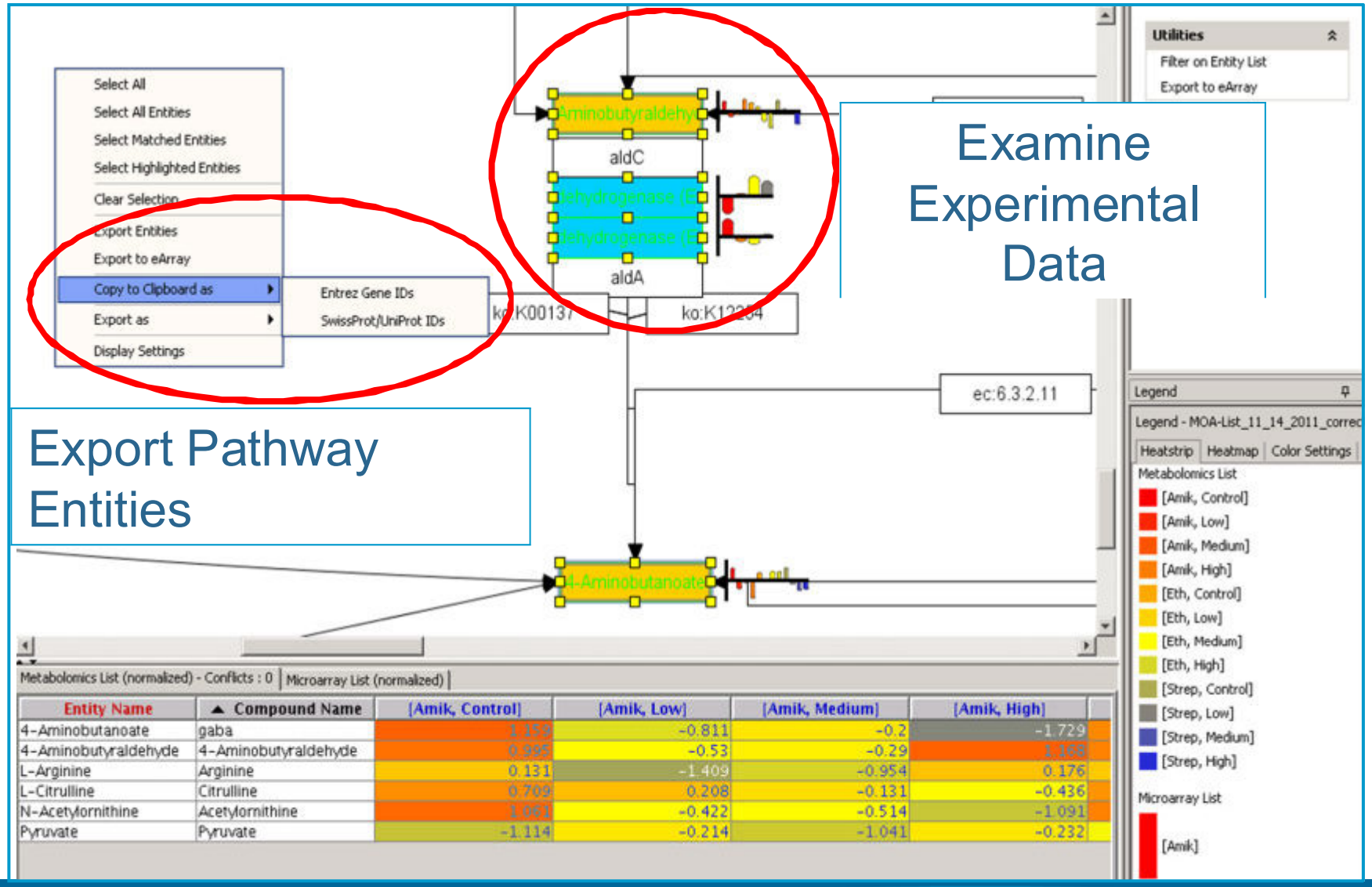
View overlaid data on WikiPathways







Examine Data And Export for Next Experiment

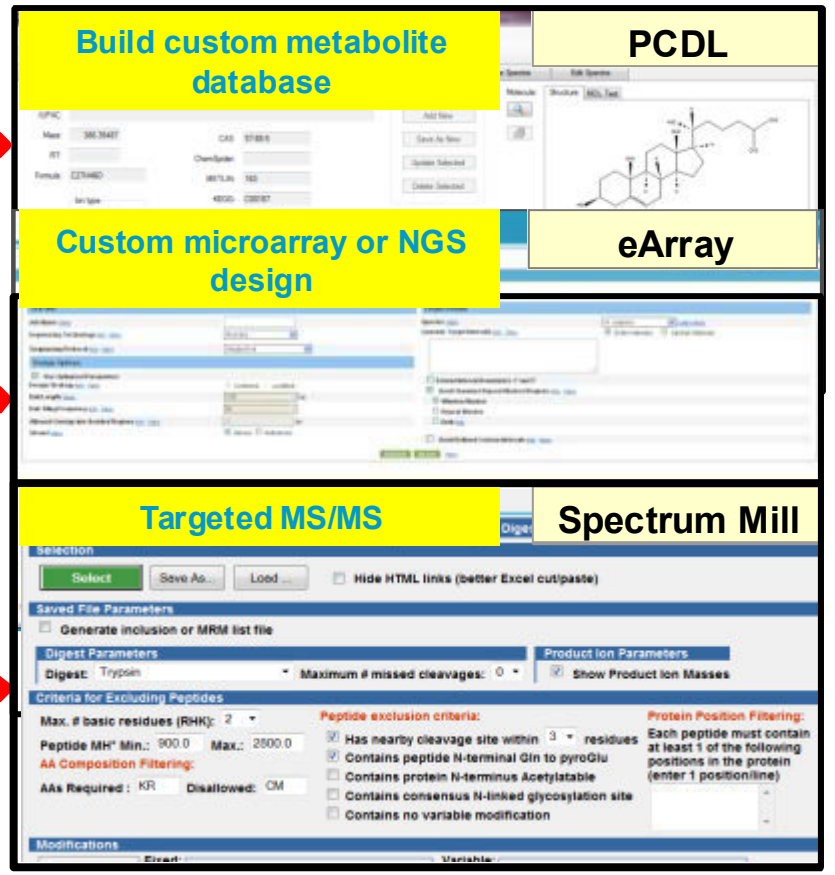


Entity Name	Compound Name	[Amik, Control]	[Amik, Low]	[Amik, Medium]	[Amik, High]
4-Aminobutanoate	gaba	1.159	-0.811	-0.2	-1.729
4-Aminobutyraldehyde	4-Aminobutyraldehyde	0.995	-0.53	-0.29	1.168
L-Arginine	Arginine	0.131	-1.409	-0.954	0.176
L-Citrulline	Citrulline	0.709	0.208	-0.131	-0.436
N-Acetylornithine	Acetylornithine	1.061	-0.422	-0.514	-1.091
Pyruvate	Pyruvate	-1.114	-0.214	-1.041	-0.232

Pathway-Directed Experiment Creation

Propose new experiments based on pathway analysis

- Re-examine acquired untargeted metabolomics data based on pathway analysis
- Design new experiments (metabolite, protein or genes) based on pathway results interpretation



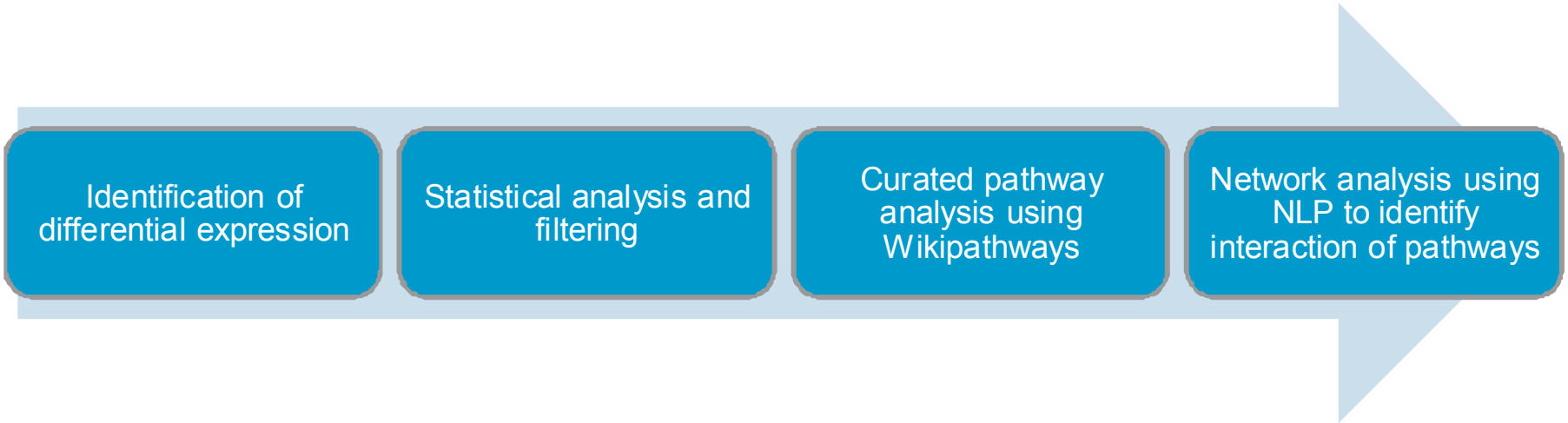
Typical workflow used for identification of a relevant pathway using GeneSpring

RESEARCH ARTICLE

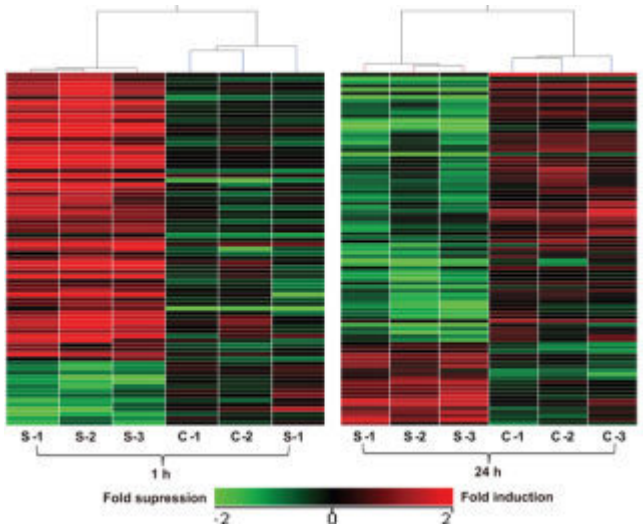
Gene Expression Changes under Cyclic Mechanical Stretching in Rat Retinal Glial (Müller) Cells

Xin Wang, Jiawen Fan, Meng Zhang, Zhongcui Sun, Gezhi Xu

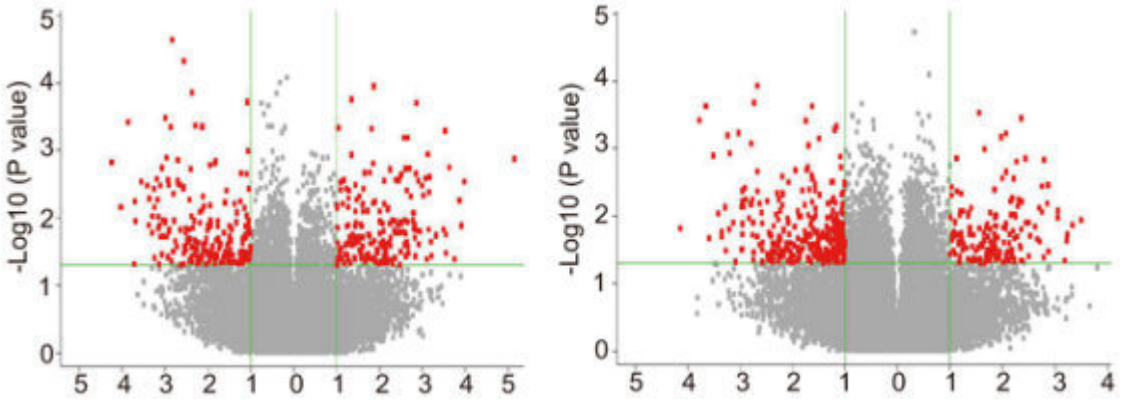
Published: May 27, 2013 • DOI: 10.1371/journal.pone.0063467



Identification of Candidate Genes



Step 1) Identification of differentially expressed genes via hierarchical cluster analysis

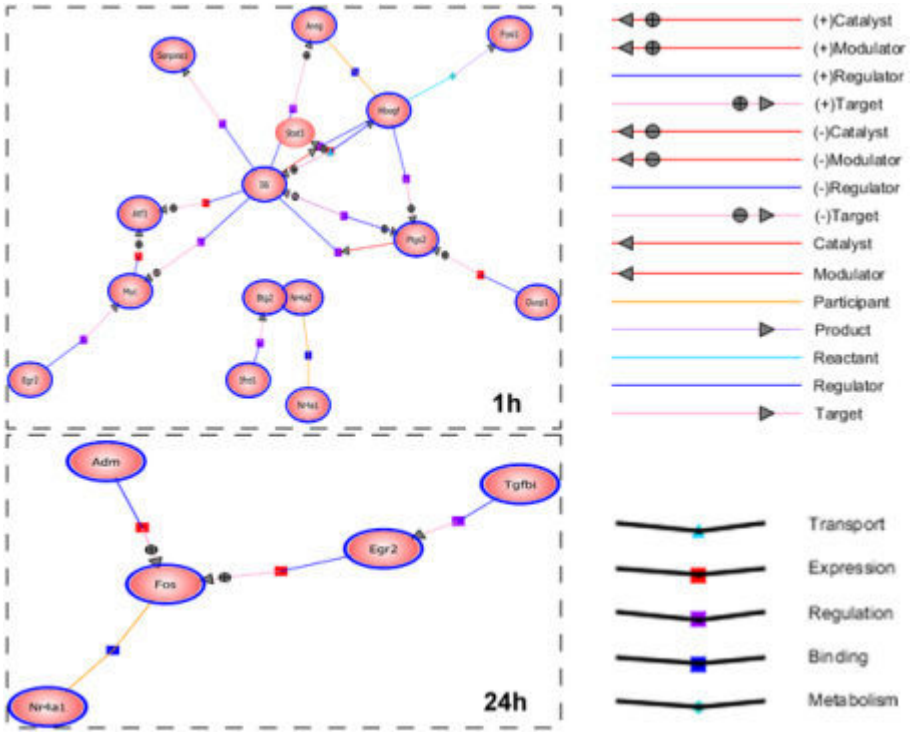


Step 2) Volcano plot of showing significantly differentially expressed between two conditions

From Differential Expression to Pathways

Pathway Name	Genes (n)	P-value
Glycerolipid metabolism	3	0.0001
Phenylalanine metabolism	2	0.0004
Nitrogen metabolism	2	0.0008
MAPK signaling pathway	4	0.0013
Tyrosine metabolism	2	0.0016
Sphingolipid metabolism	2	0.0026
Retinol metabolism	2	0.0066
B cell receptor signaling pathway	2	0.0089
Metabolic pathways	6	0.0171
Sulfur metabolism	1	0.0203
Vascular smooth muscle contraction	2	0.0209
Axon guidance	2	0.0227
Other glycan degradation	1	0.0286
beta-Alanine metabolism	1	0.0385
Nicotinate and nicotinamide metabolism	1	0.0401
Pentose and glucuronate interconversions	1	0.0401
Histidine metabolism	1	0.0418
Calcium signaling pathway	2	0.0431
Ascorbate and aldarate metabolism	1	0.0434
Homologous recombination	1	0.0450

doi:10.1371/journal.pone.0063467.t004



Step 3) Significantly changed pathways in Müller cells identified using pathway analysis in GeneSpring

Step 4) The Protein-Protein Interactions analysis was further performed to identify the direct interaction of these genes products in GeneSpring

Pathway analysis showed significant changes in MAPK signaling at both conditions.
Network analysis shows interaction of MAPK with other gene products.
Compare network analysis/extension in Cytoscape.

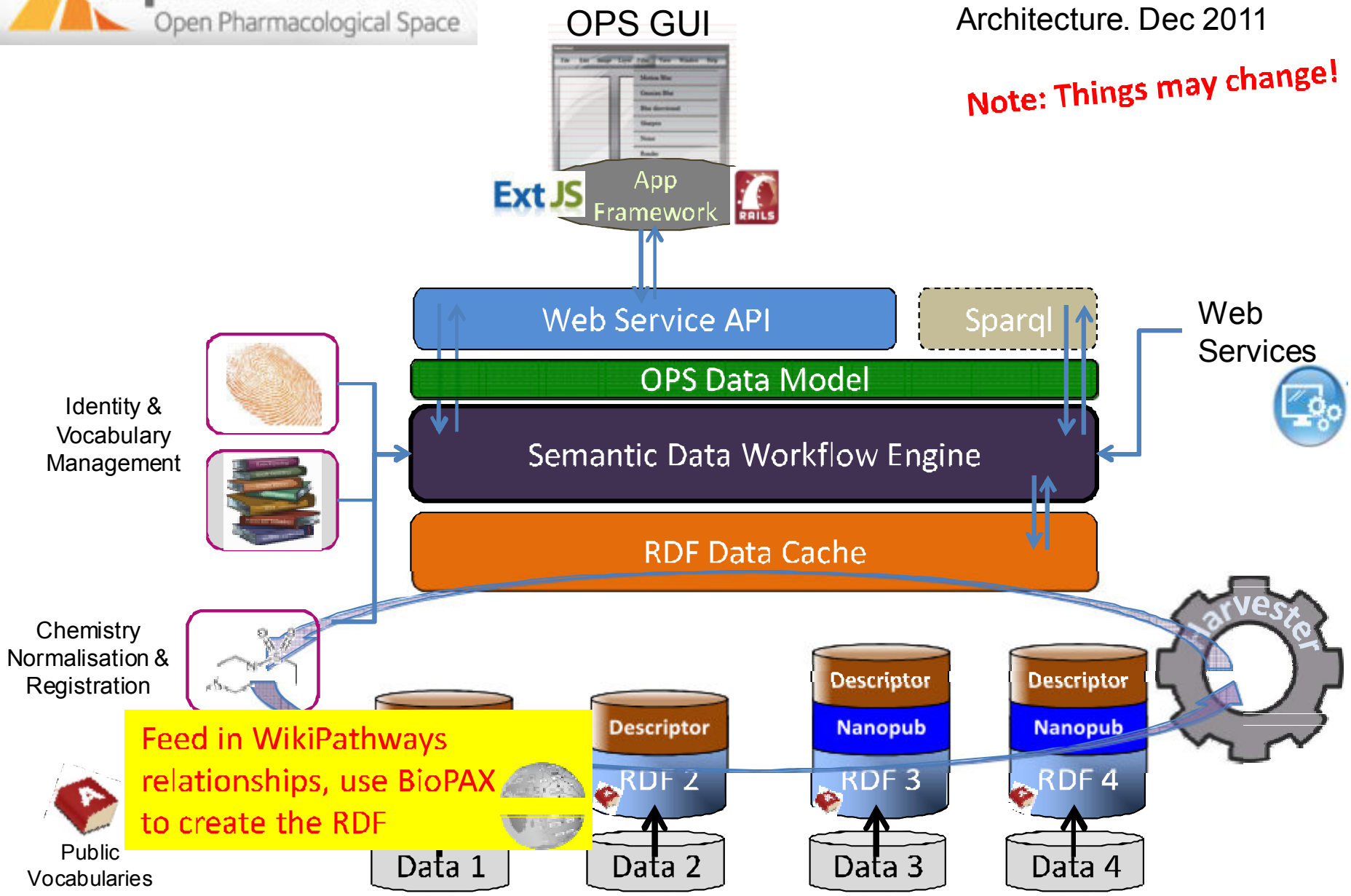
Combine further with

- **Open Knowledge**
  e.g. IMI semantic web project Open PHACTS
  Pathway content and extension
- **Open Data**
  e.g. ISAtab based study capturing
  in phenotype database (dbNP)
  pathway analysis and profiling

Open PHACTS
Open Pharmacological Space

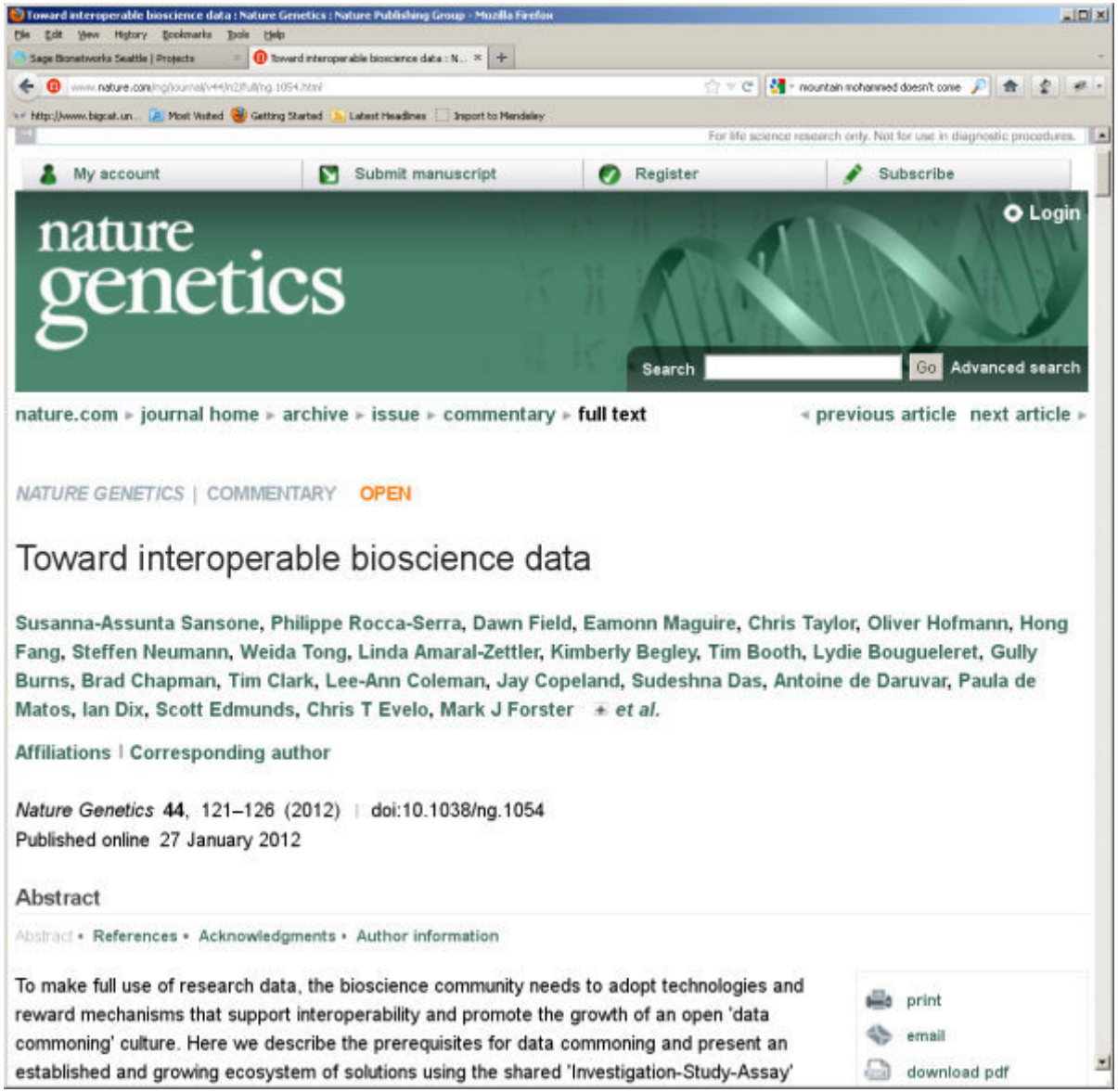
OPS Framework Architecture. Dec 2011

Note: Things may change!

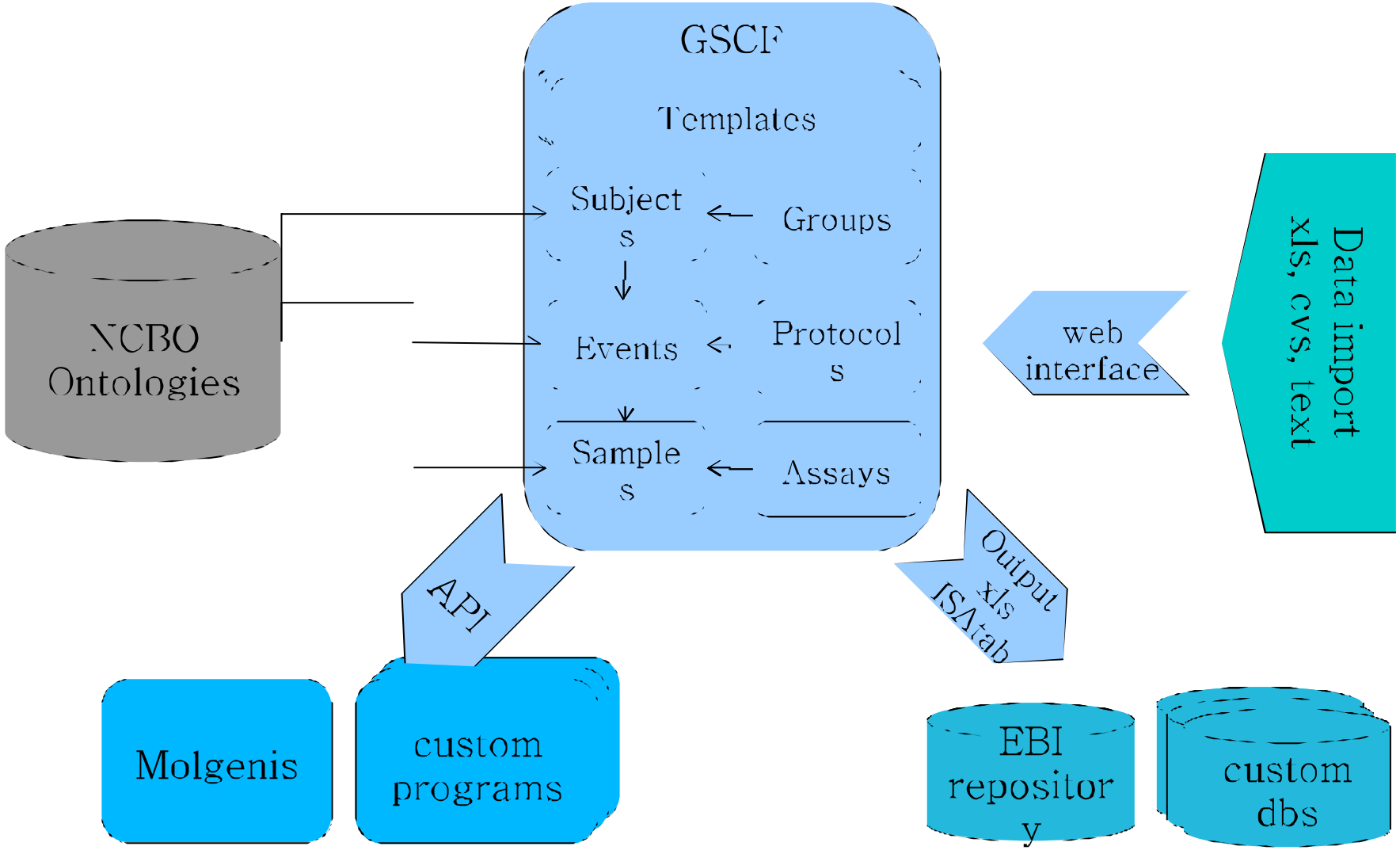


Data capturing

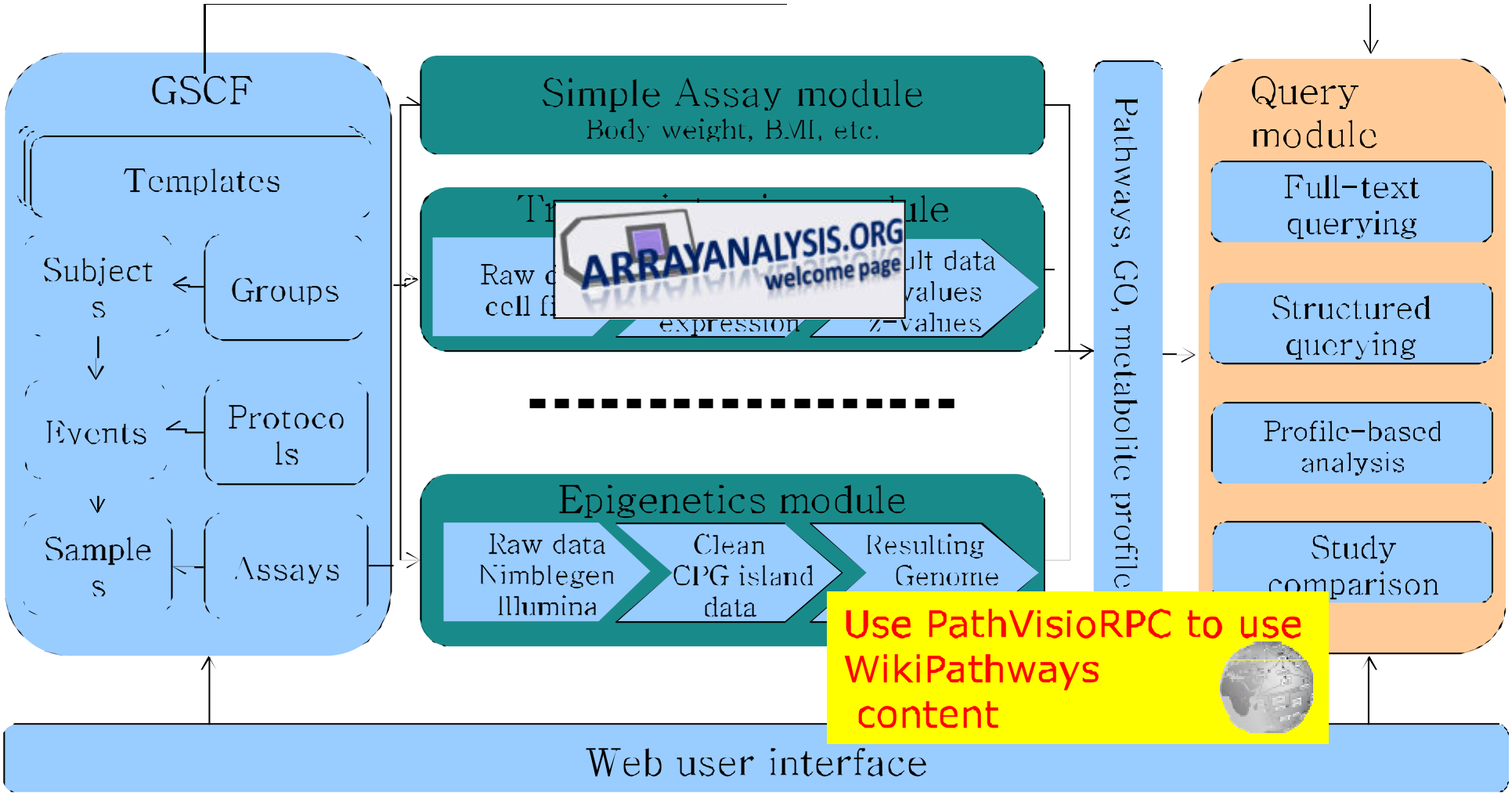
Using ISA to connect to the rest of world



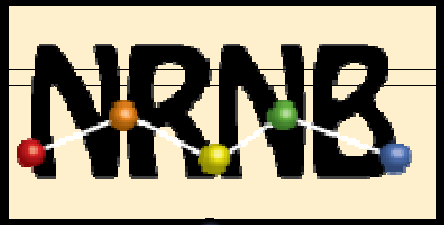
Generic Study Capture Framework
Data input / output



dbNP Architecture



Thomas Kelder
Martijn van Iersel
Kristina Hanspers
Martina Kutmon
Andra Waagmeester
Chris Evelo
Bruce Conklin



nrnb.org

wikipathways.org

Acknowledgements