



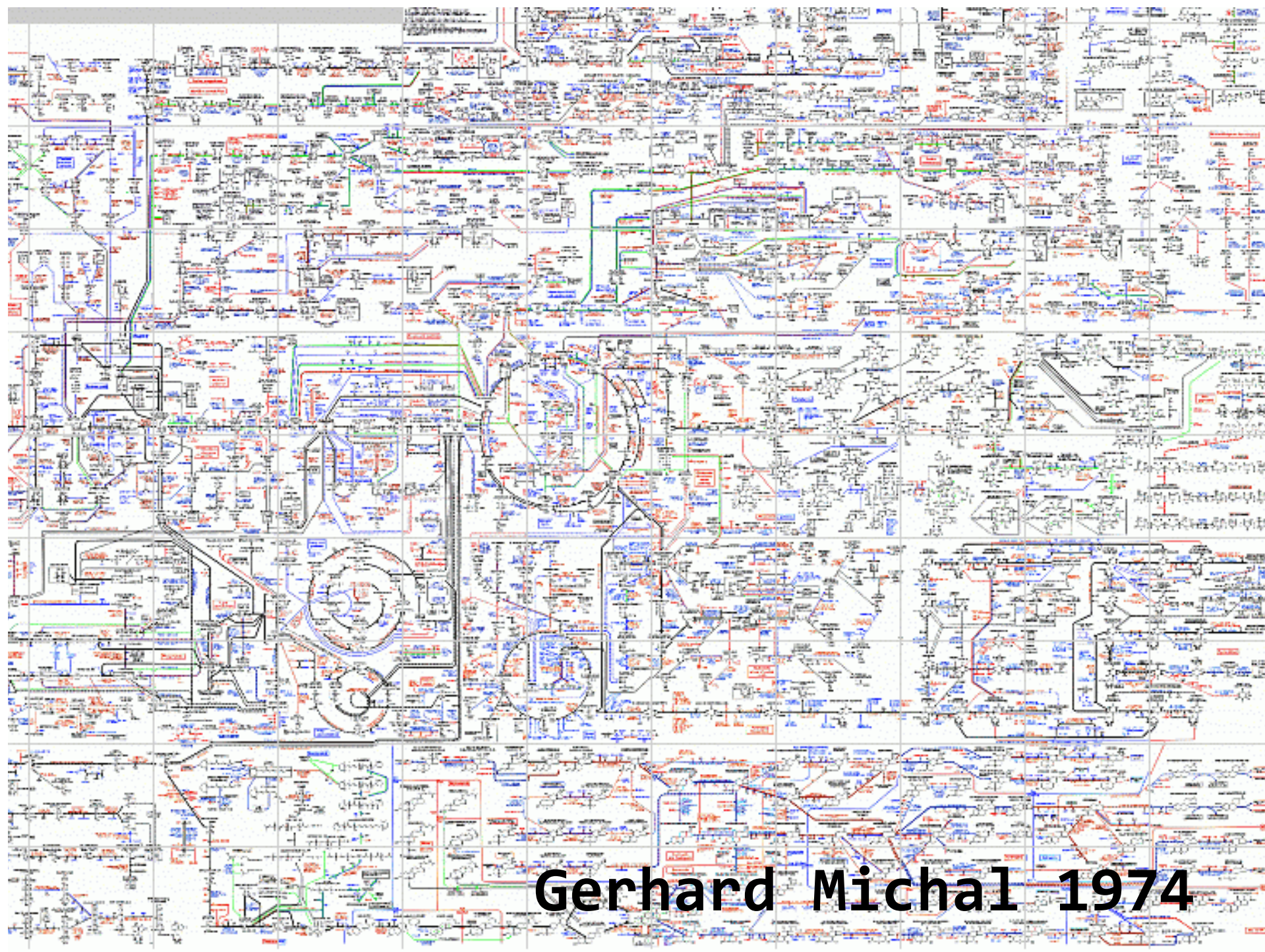
# WikiPathways: how open source code and open data can make omics technology more useful

@Chris\_Evelo

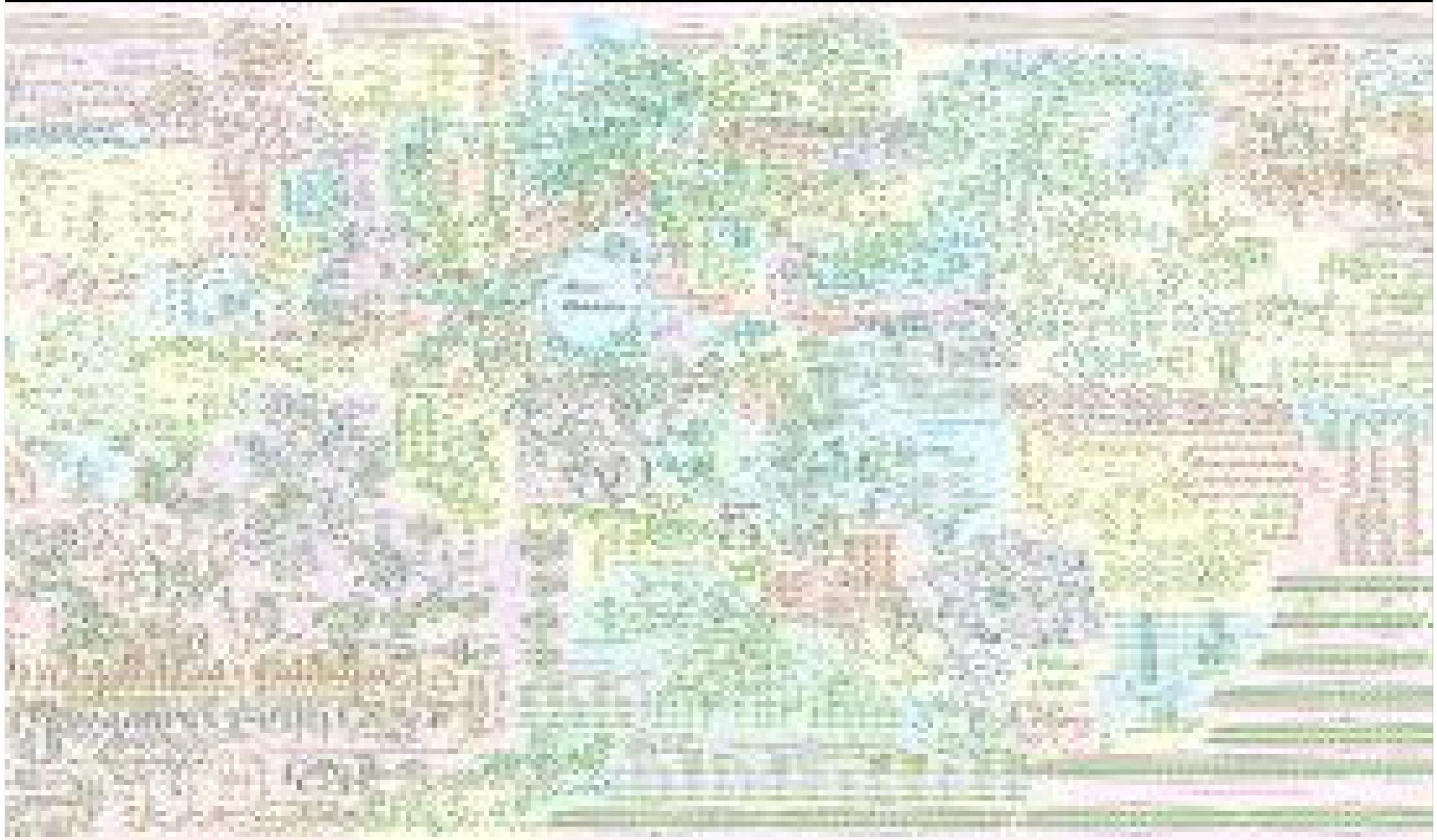
Department of Bioinformatics - BiGCaT  
Maastricht University





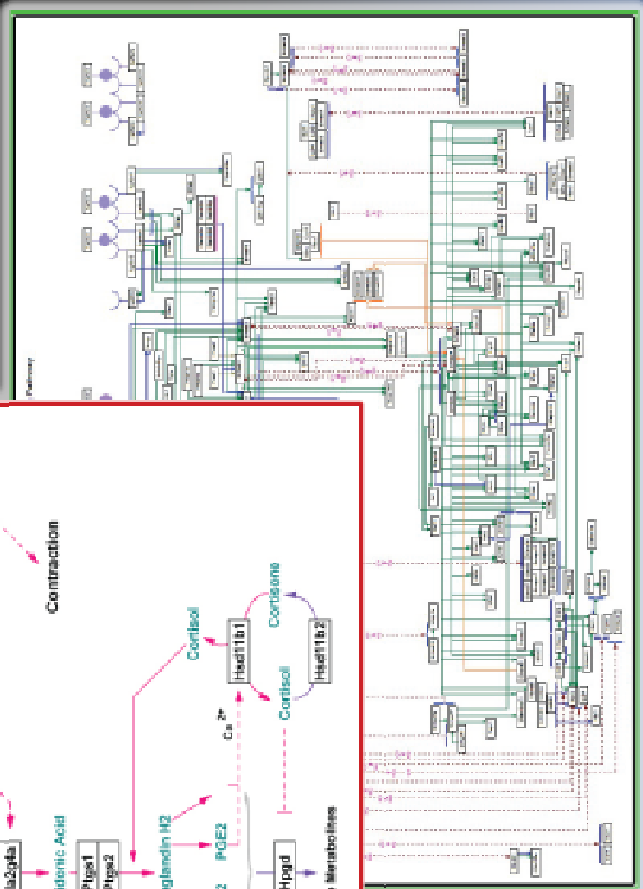
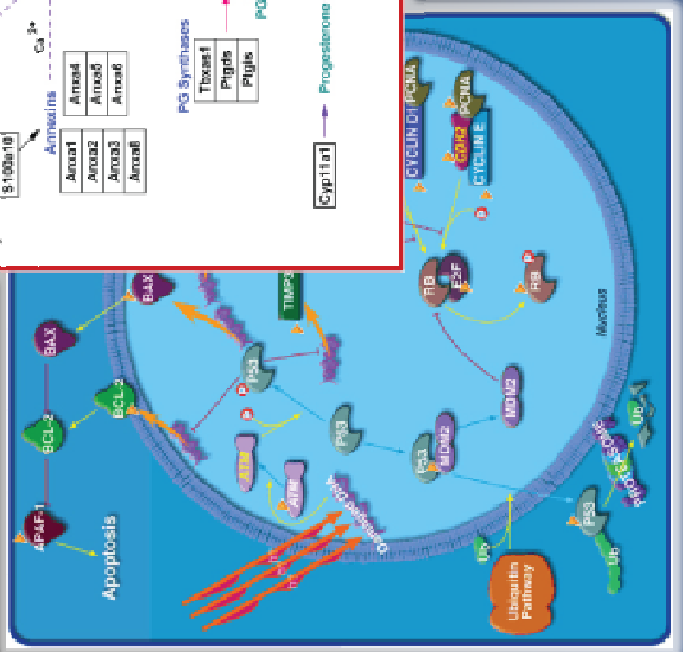
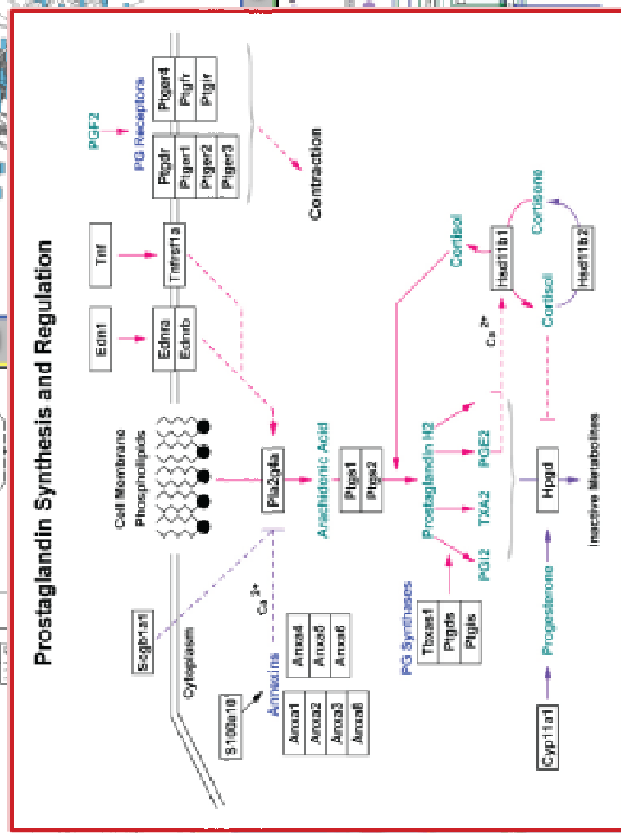
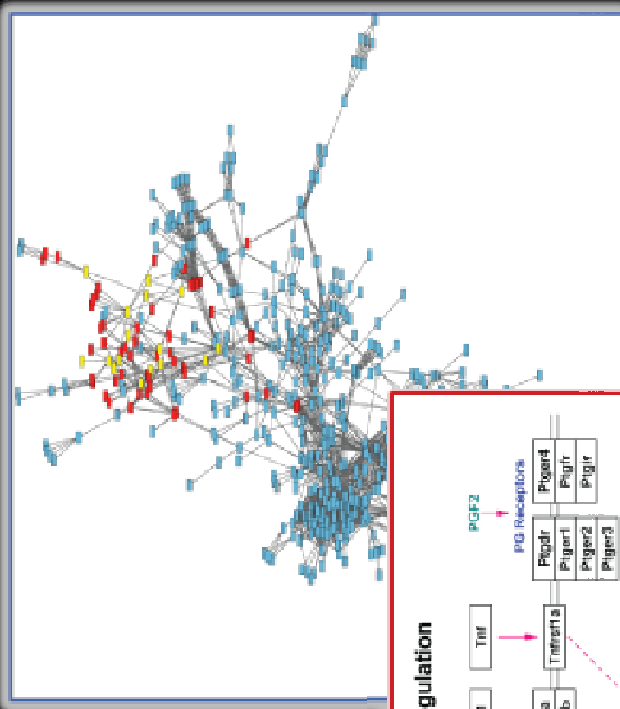
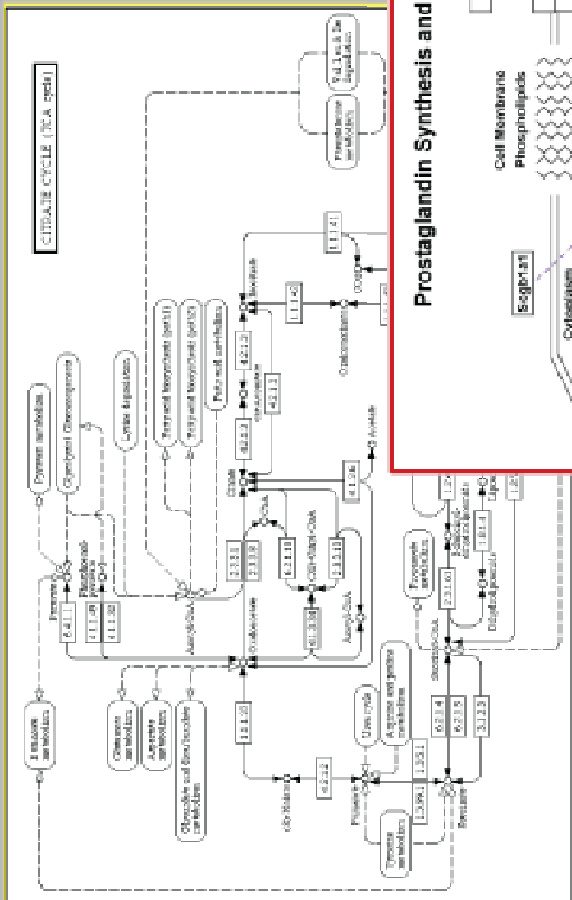


# Recon2 a Google map of human metabolism

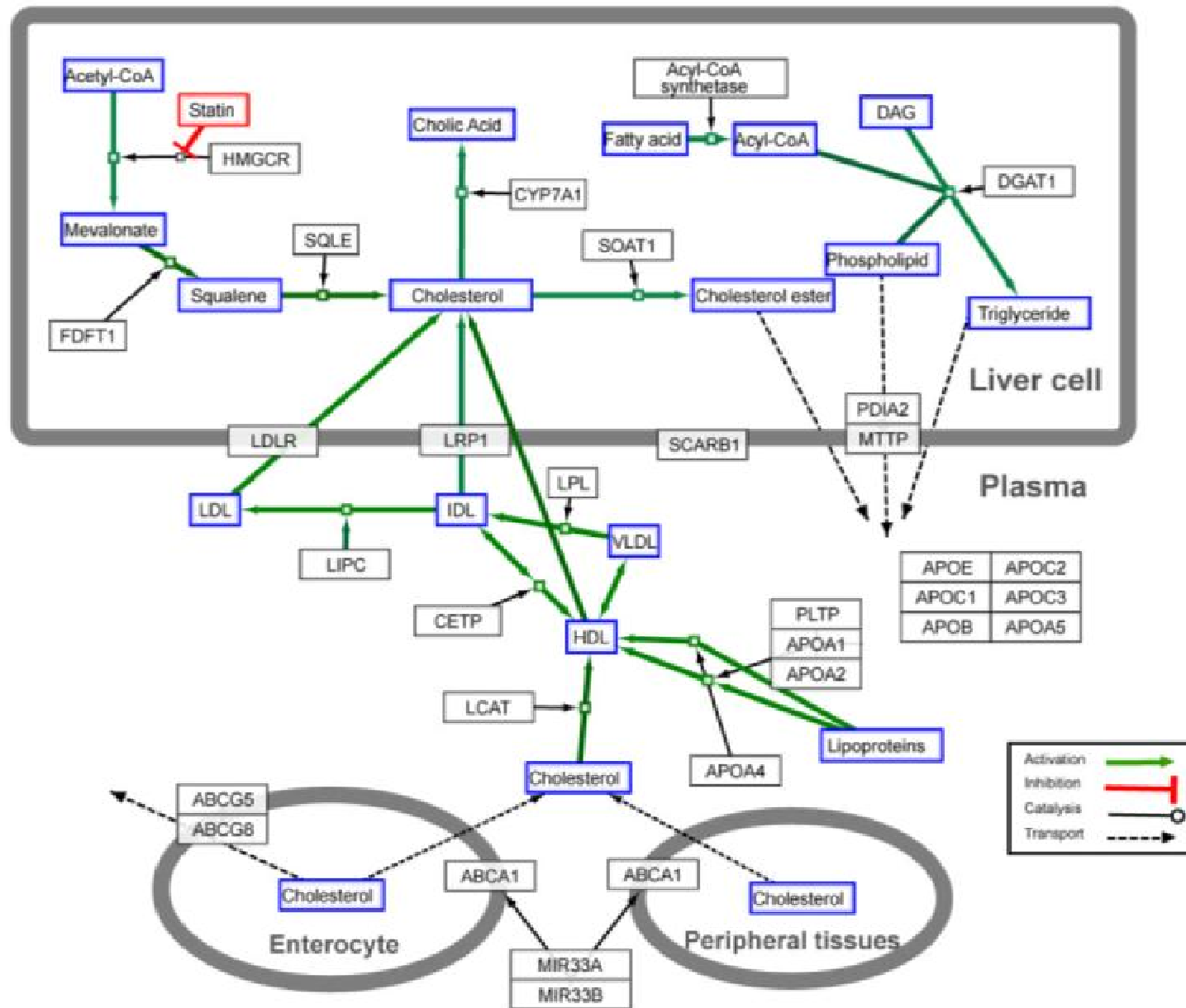


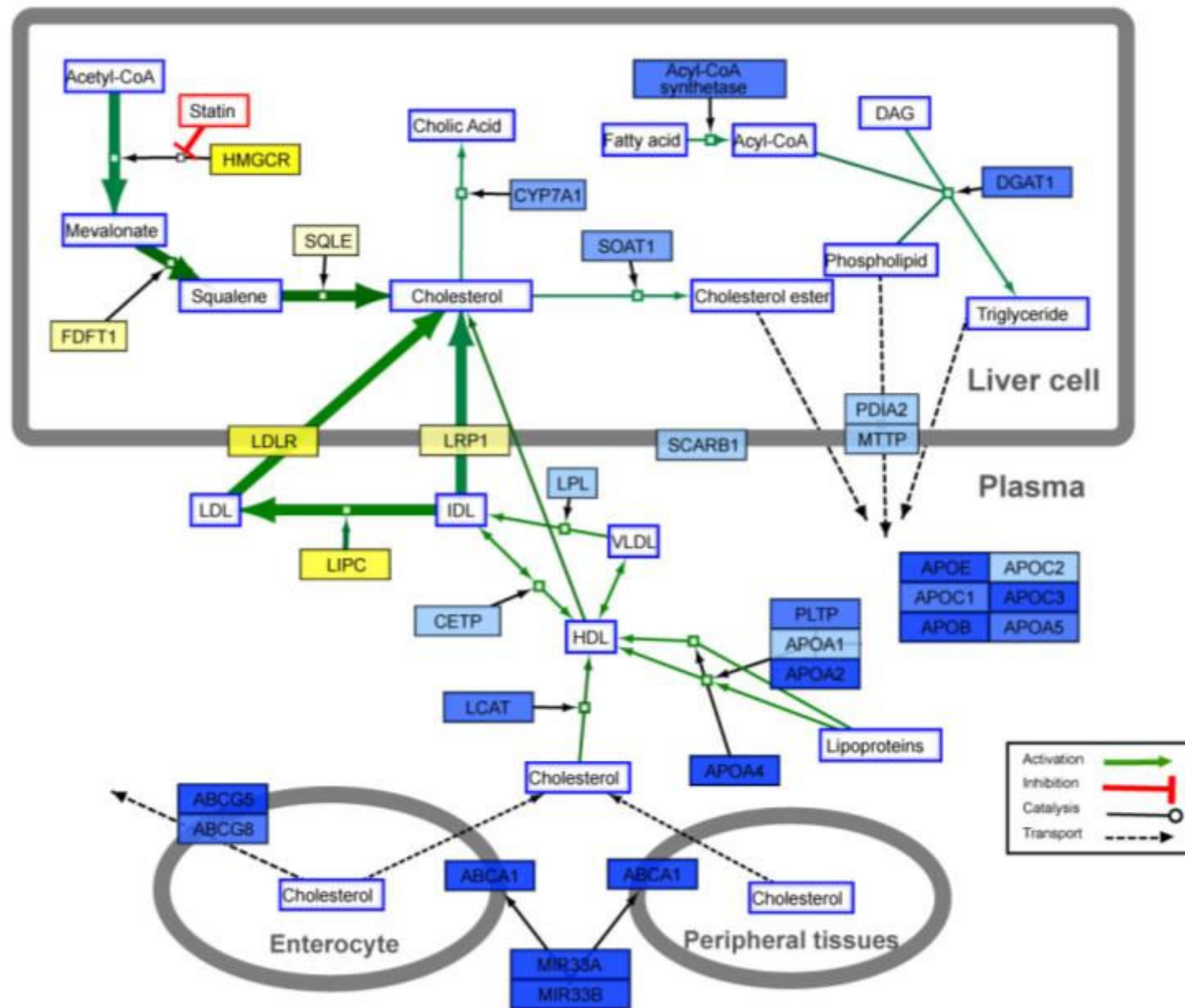
Collaborating Systems Biology and Metabolomics groups: 2013

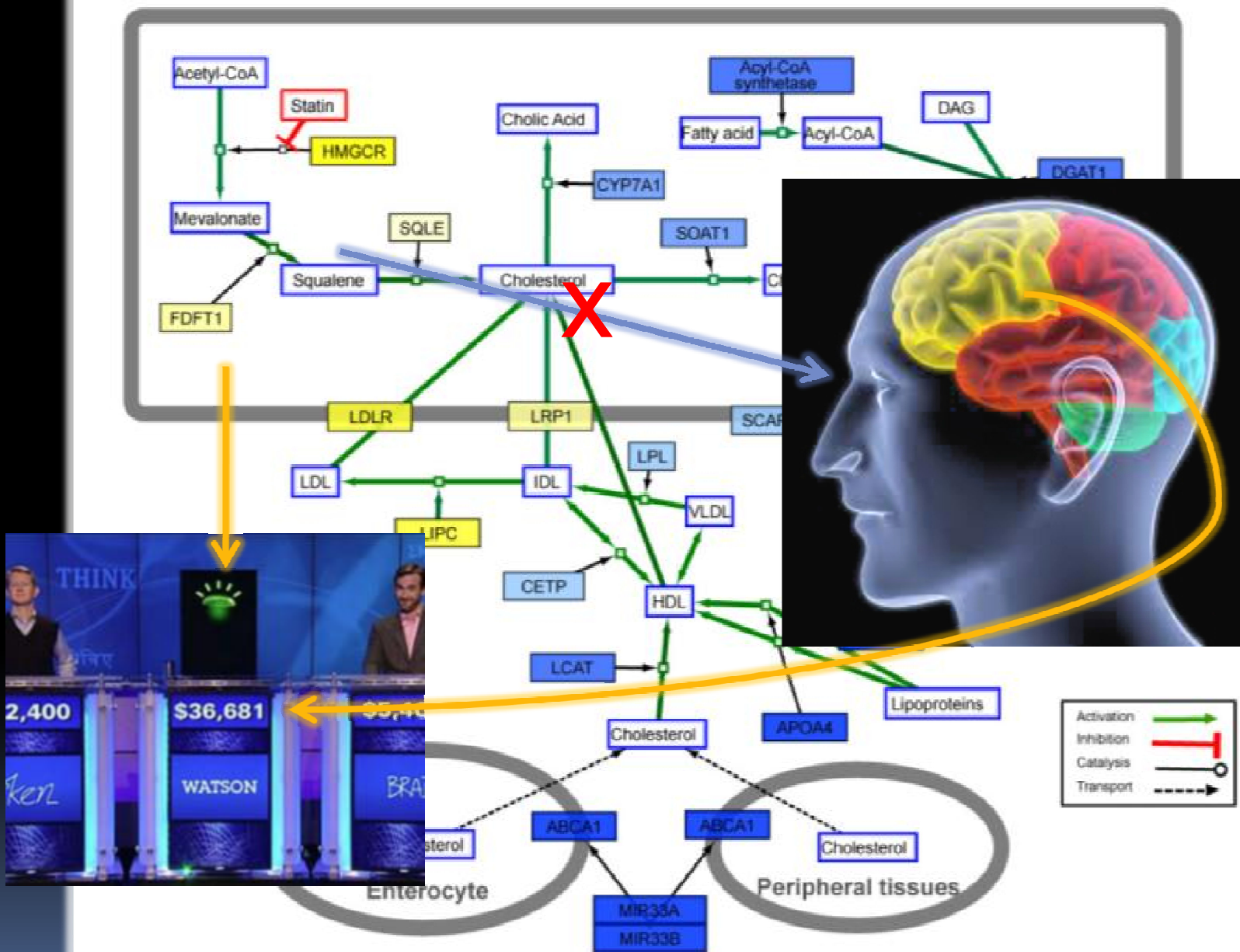




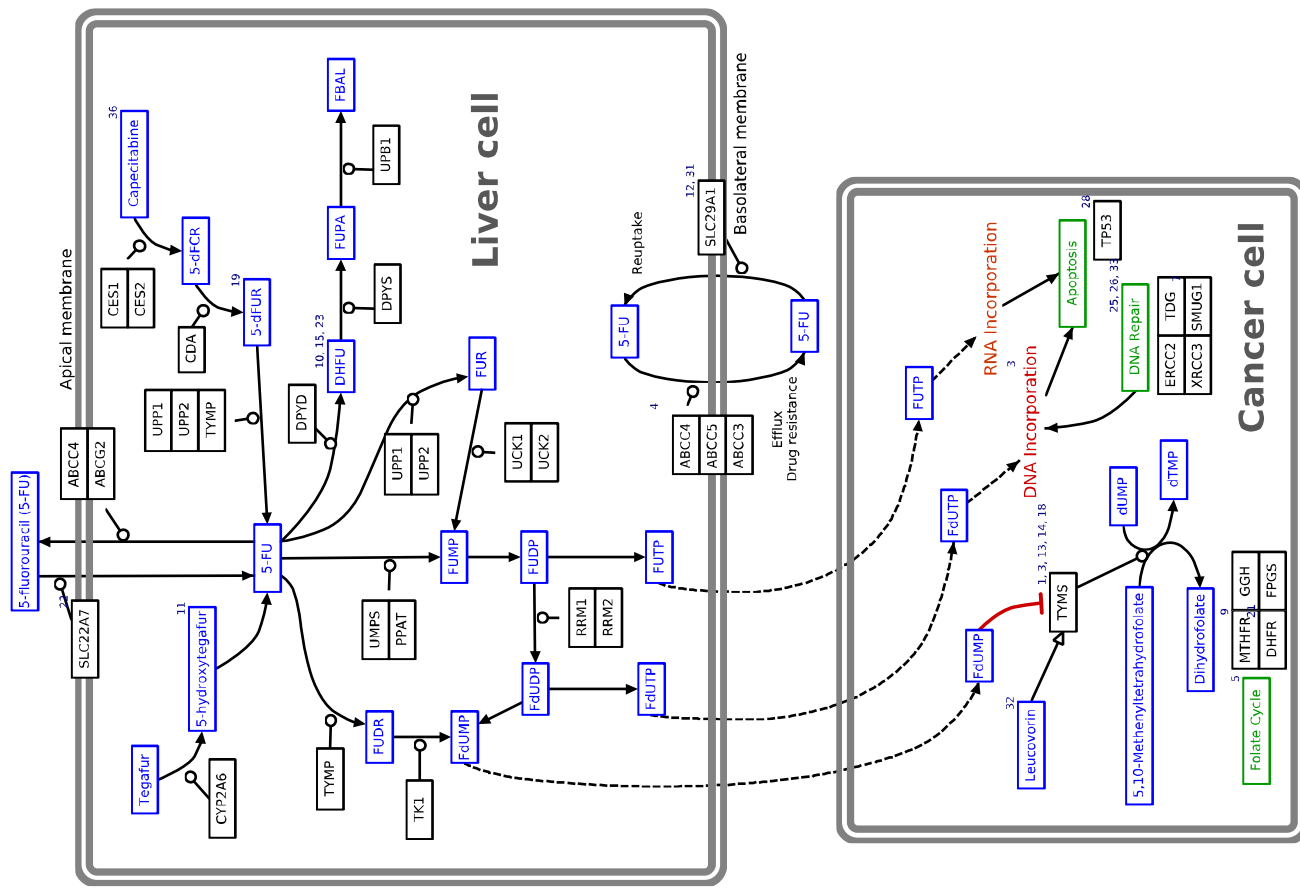








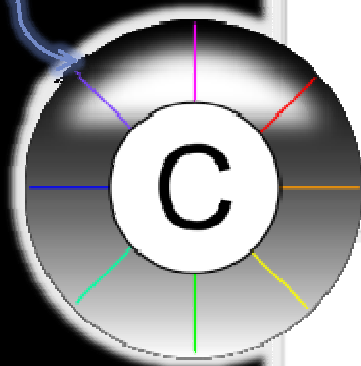






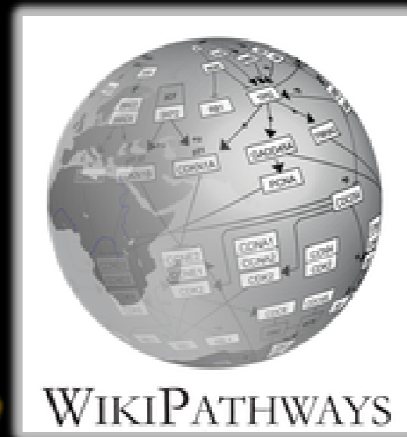






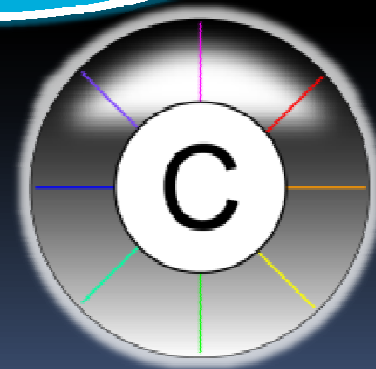
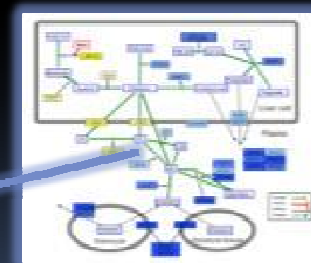


**data**



**ideas**

**synthesis**

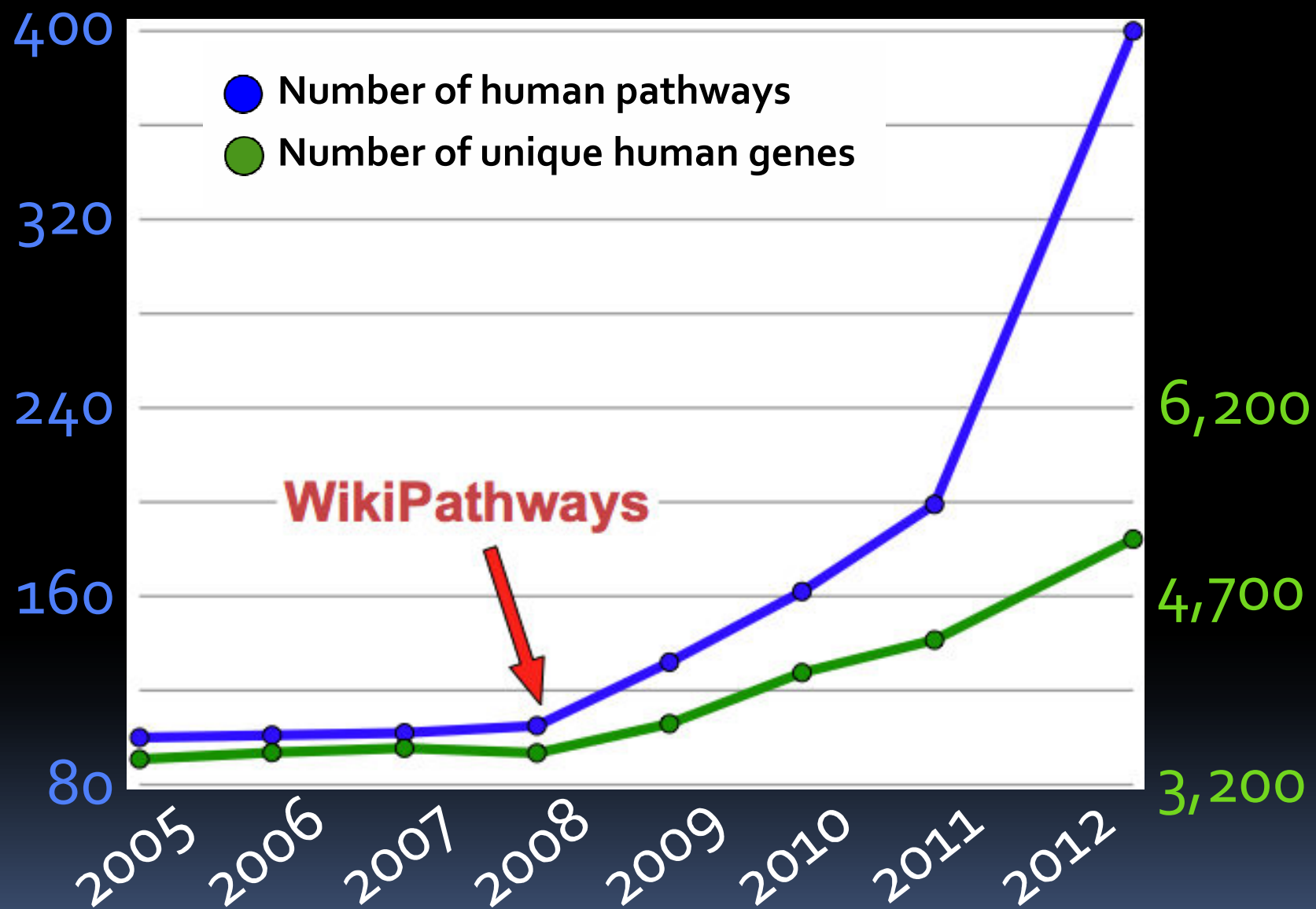


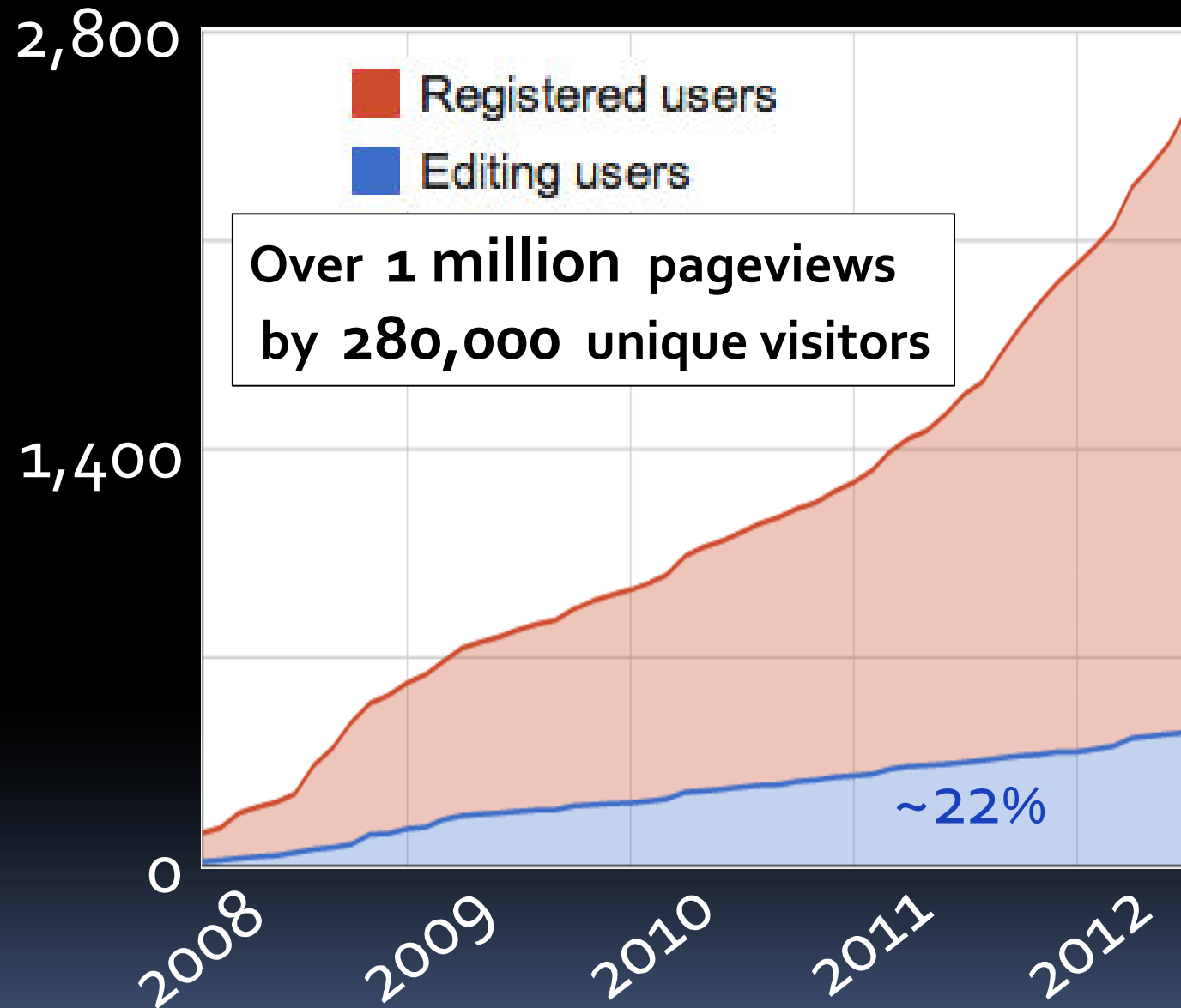
# Set Early Milestones

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









# 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*



# PathVisio

a tool to edit and analyze biological pathways

Home Getting Started ▾ Support/Help ▾ Downloads ▾ Plugins ▾ Cite Us About ▾

Search

## What is PathVisio?

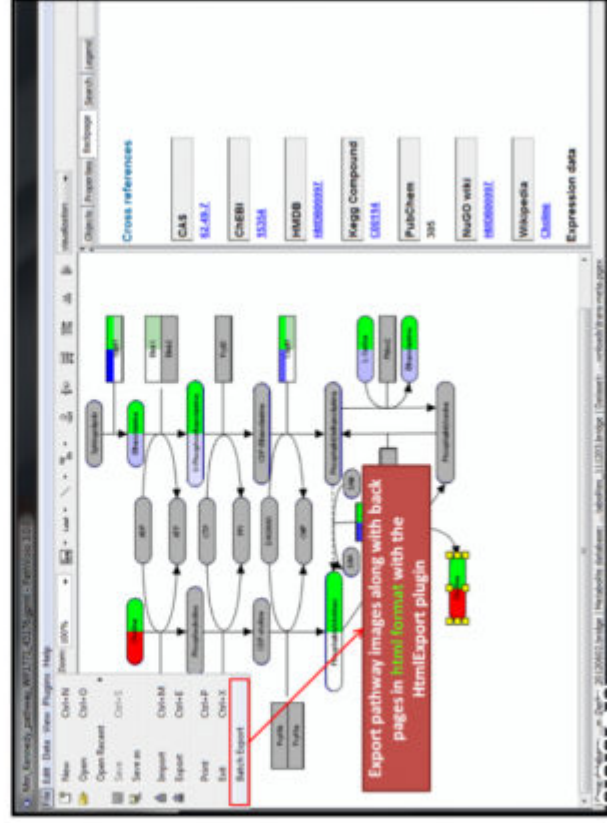
PathVisio is a free open-source biological pathway analysis software that allows you to draw, edit and analyze biological pathways. [Learn more.](#)

## 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



[http://www.agilent.com/univ\\_relation/TLP/index.shtml](http://www.agilent.com/univ_relation/TLP/index.shtml)

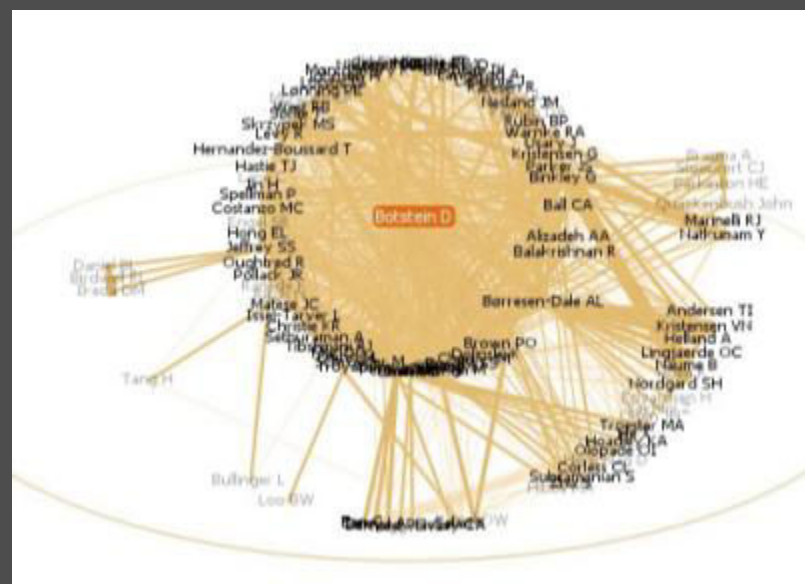
## Thought Leader Awards

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

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

- Candidates selected based on scientific leadership, productivity, project significance (invitational program)
- High-level executive sponsorship and active support throughout Agilent enable breakthrough research

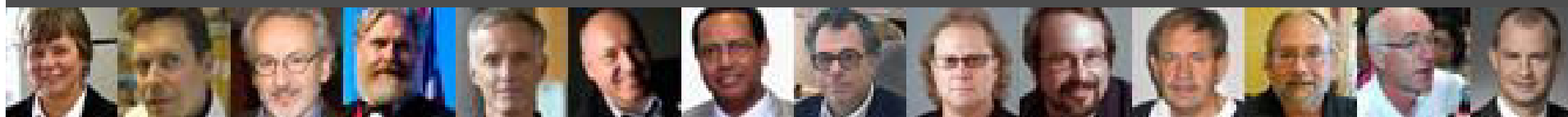
# TL Network



## Early Career Professor Award

Establish strong collaborative relationships with highly promising early career professors

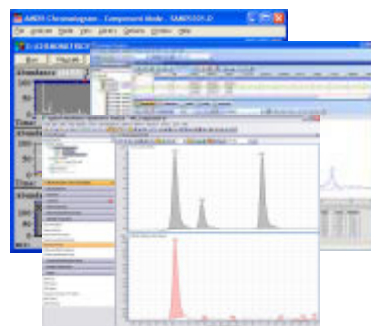
2013 focus:  
Contributions to cancer diagnostics



# Agilent Integrated Biology Workflows



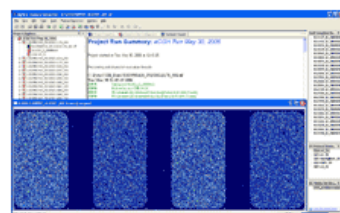
LC/MS  
GC/MS



MassHunter Qual/Quant  
ChemStation AMDIS



Microarrays



Feature Extraction



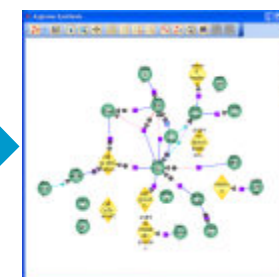
NGS



Alignment to Reference Genome



GeneSpring Platform



Biological Pathways



Agilent Technologies



# GeneSpring

*Agilent's Platform for Multi-Omics Data*

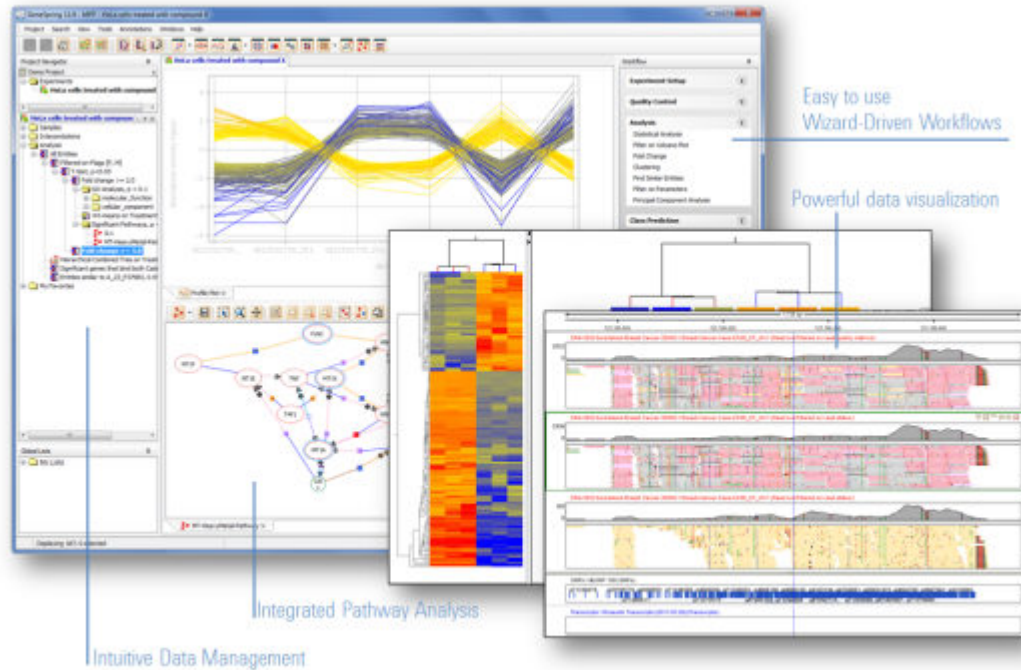
10 More than 100 Hundreds of 10000s 100000 References  
Years of Experience Technology Types Genes Proteins In Google Scholar



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)



Agilent Technologies

# GeneSpring modules

## GX

mRNA

microRNA

QPCR

Alternative Splicing

GWAS & CNV via SNP arrays

## GX (Gene Expression)

## NGS (Next-Gen Sequencing)

SureSelect Target Enrichment

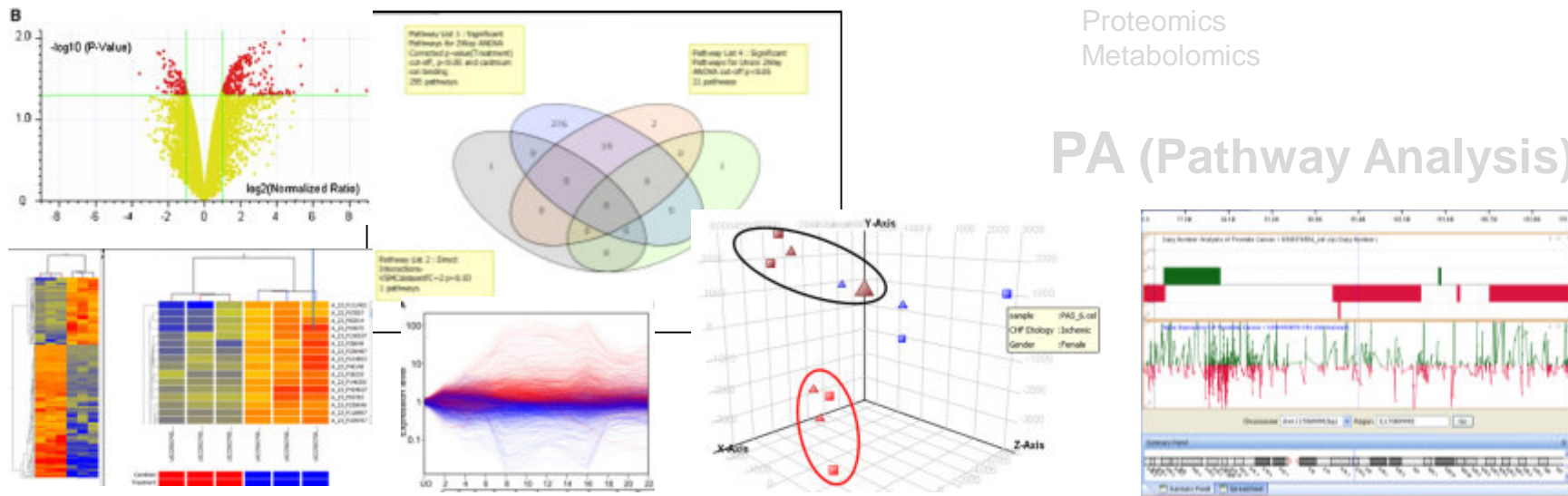
Whole Genome Sequencing

## MPP (Mass Profiler Pro)

Proteomics

Metabolomics

## PA (Pathway Analysis)



Agilent Technologies

# GeneSpring modules

## NGS

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

### 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)

**GX** (Gene Expression)

## NGS (Next-Gen Sequencing)

SureSelect Target Enrichment  
Whole Genome Sequencing

**MPP** (Mass Profiler Pro)

Proteomics  
Metabolomics

**PA** (Pathway Analysis)



Agilent Technologies

# 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)



Agilent Technologies

# GeneSpring modules

## PA

# Multi-Omic Analysis

## Canonical Pathways

## Network Discovery

## GX (Gene Expression)

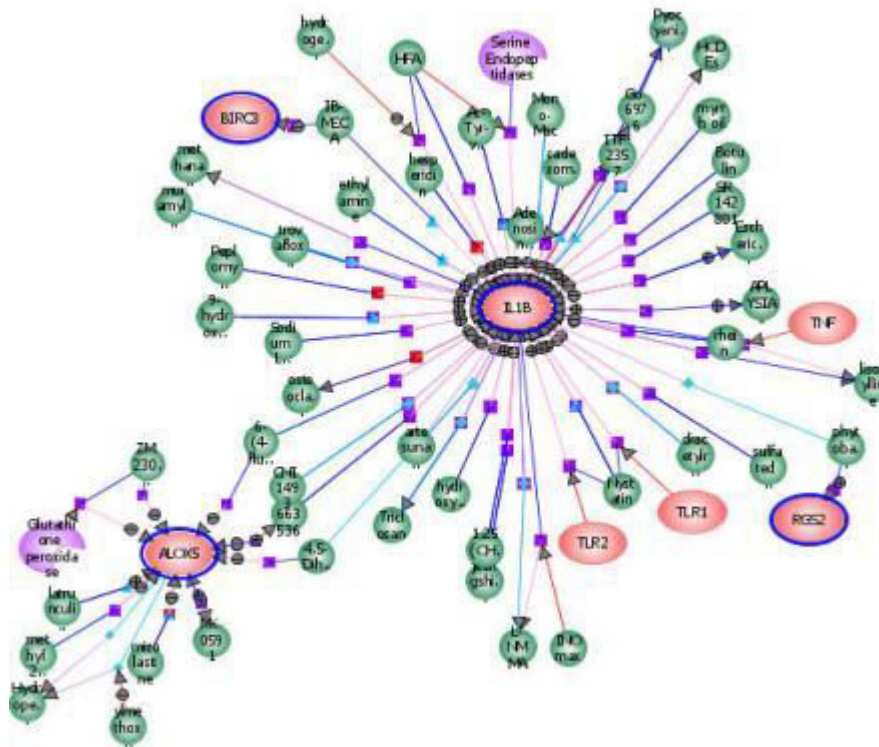
# NGS (Next-Gen Sequencing)

SureSelect Target Enrichment  
Whole Genome Sequencing

# MPP (Mass Profiler Pro)

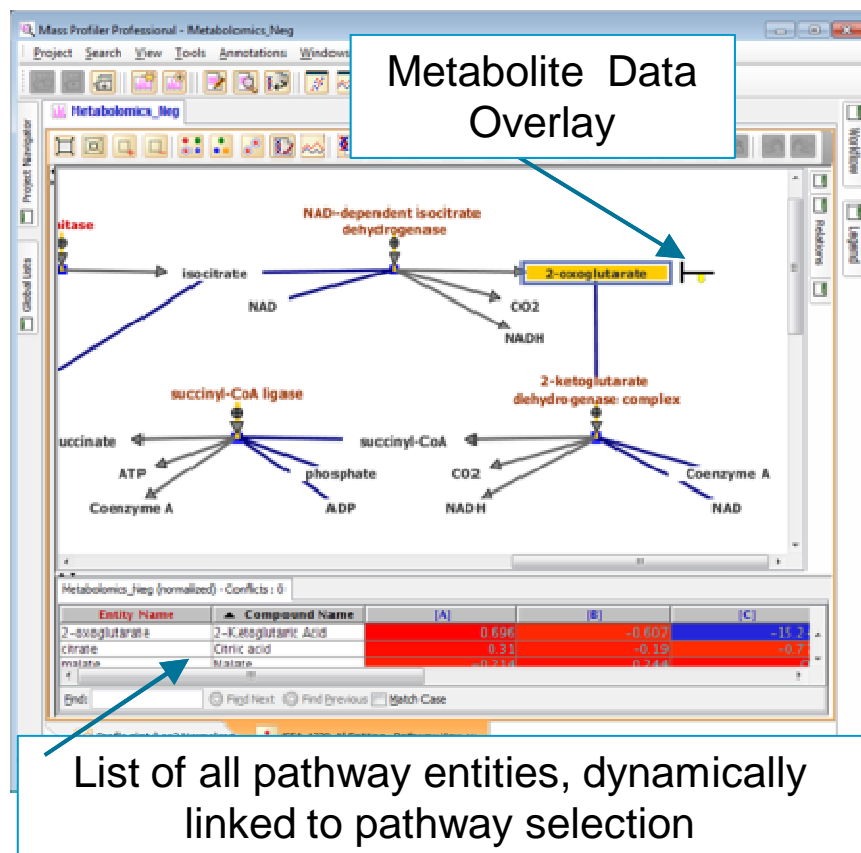
Proteomics  
Metabolomics

## 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:

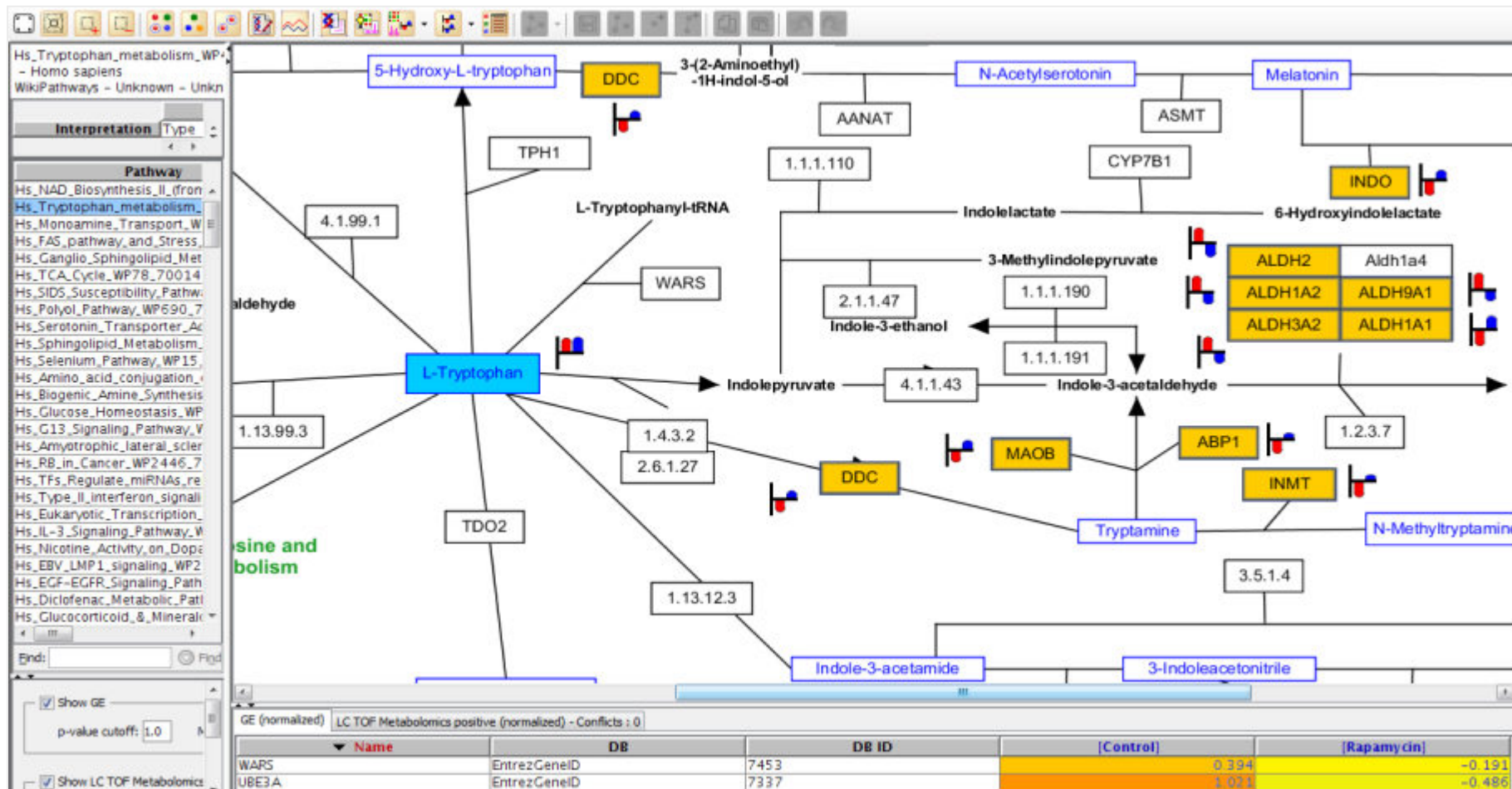
- WikiPathways
- BioCyc

- 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



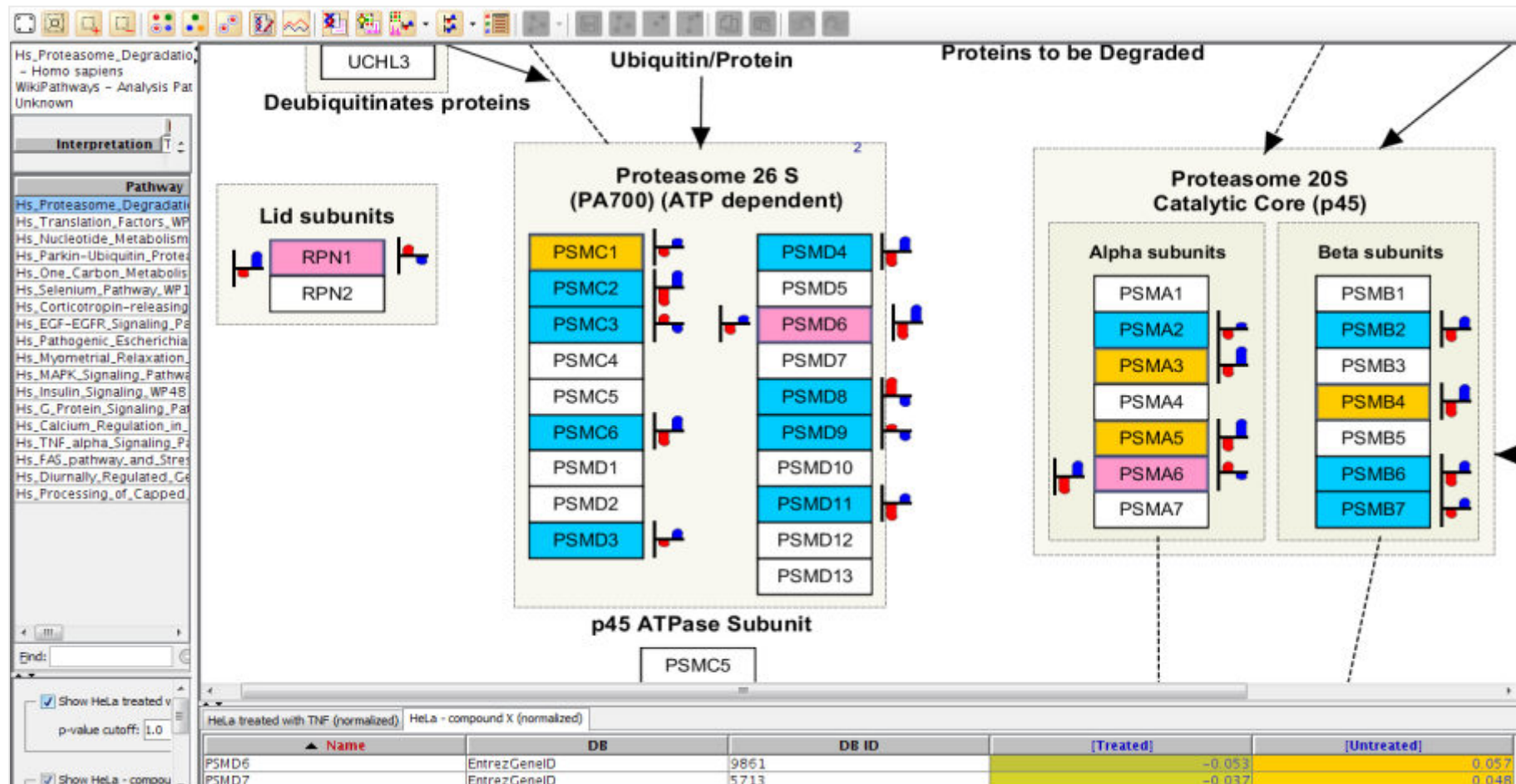
metabolites

proteins



Agilent Technologies

# Genomics and Proteomics data overlay on Proteasome Degradation WikiPathway



Proteomics

Genomics

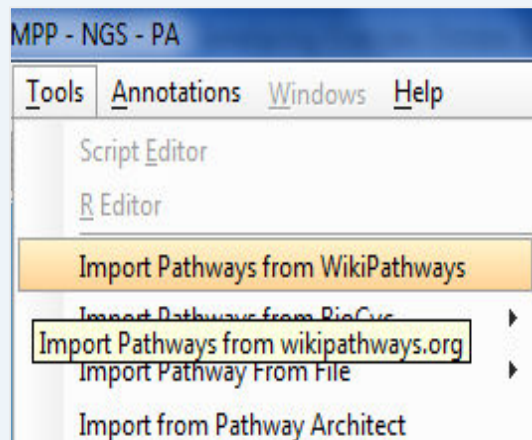
Proteomics and Genomics



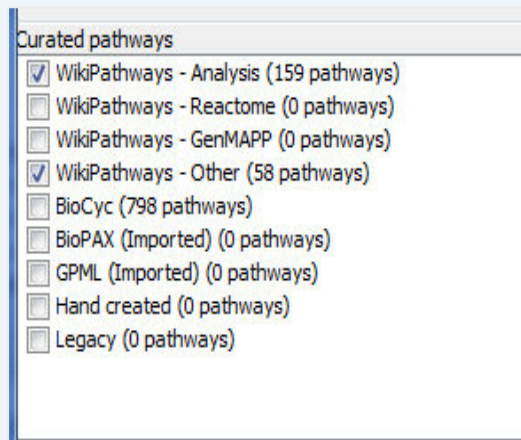
Agilent Technologies

# Analysis with WikiPathways in Pathway Architect

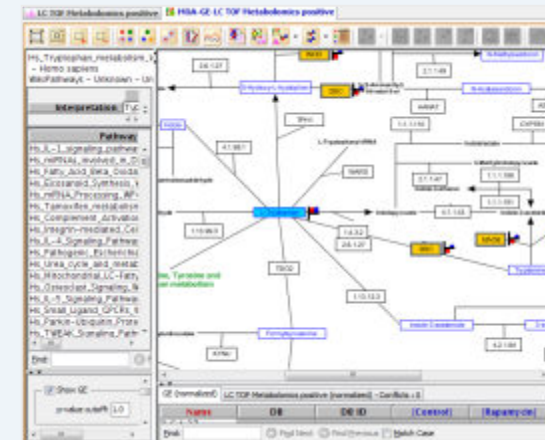
Import of WikiPathways



Select WikiPathways for analysis



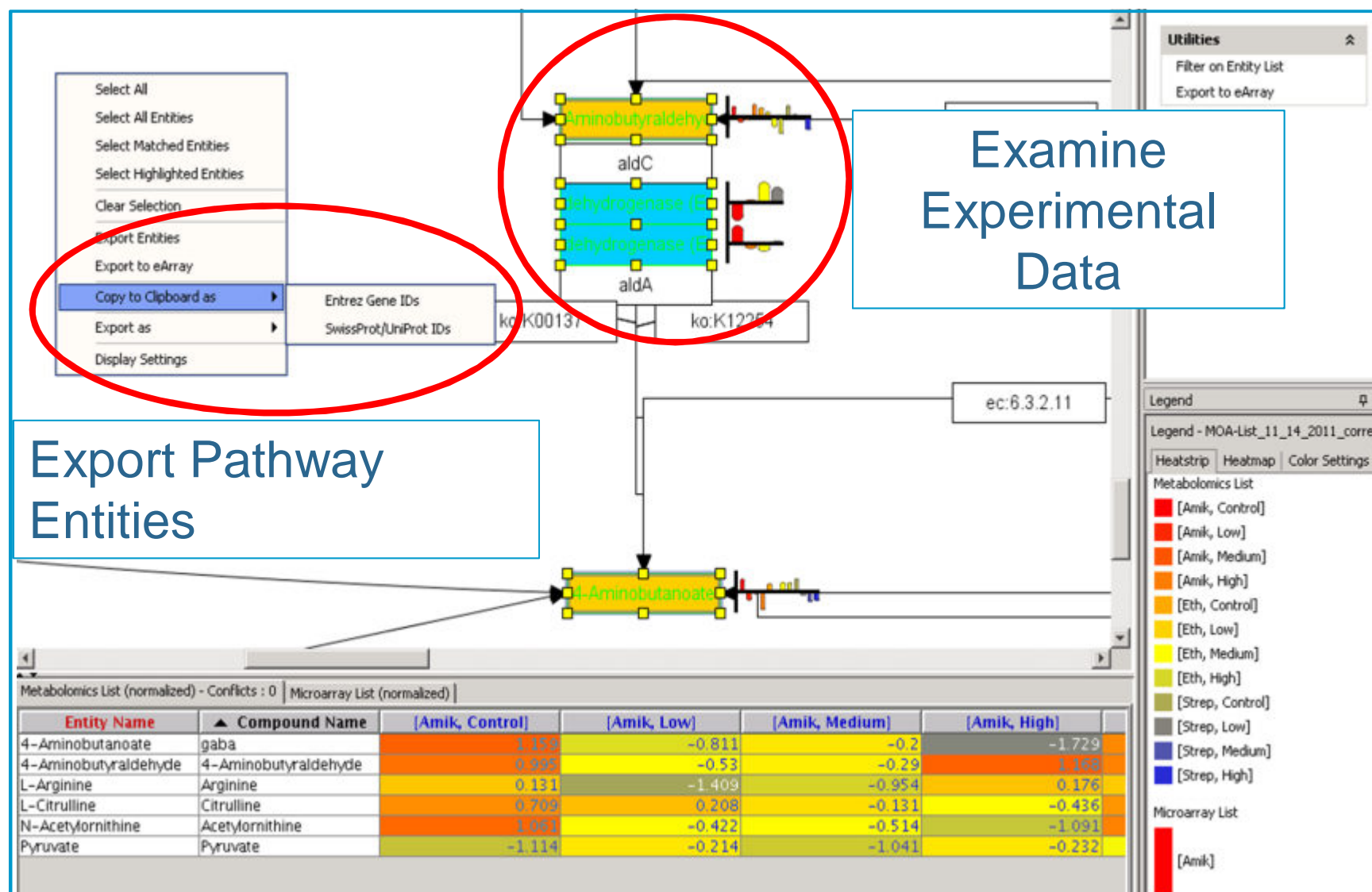
View overlaid data on WikiPathways



Agilent Technologies



# Examine Data And Export for Next Experiment





# 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



**Build custom metabolite database**

**PCDL**

**Custom microarray or NGS design**

**eArray**

**Targeted MS/MS**

**Spectrum Mill**

The right side of the slide displays four software interfaces for experiment creation. The top two are 'Build custom metabolite database' and 'PCDL'. The middle two are 'Custom microarray or NGS design' and 'eArray'. The bottom two are 'Targeted MS/MS' and 'Spectrum Mill'. A red arrow points from the 'Propose new experiments based on pathway analysis' text to the 'Targeted MS/MS' interface.



# 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  
differential expression

Statistical analysis and  
filtering

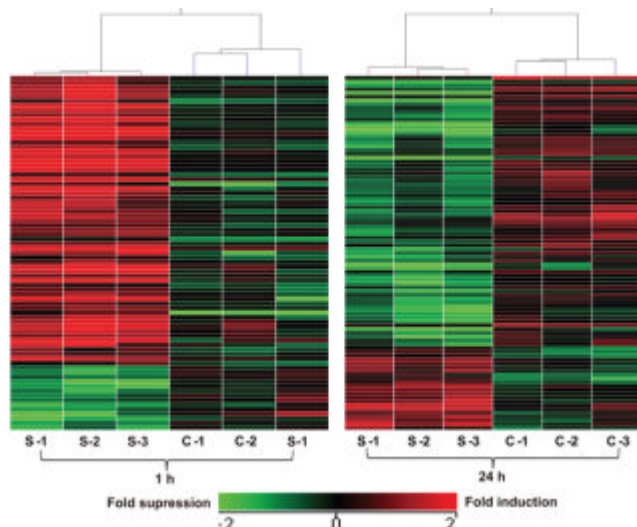
Curated pathway  
analysis using  
Wikipathways

Network analysis using  
NLP to identify  
interaction of pathways

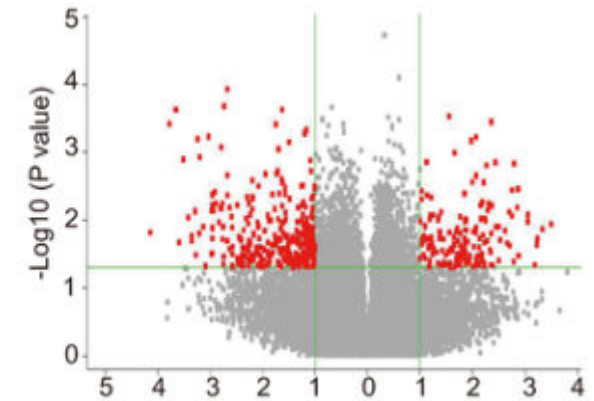
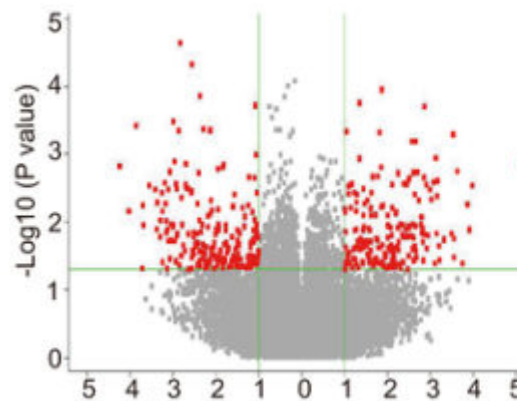


Agilent Technologies

# 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**

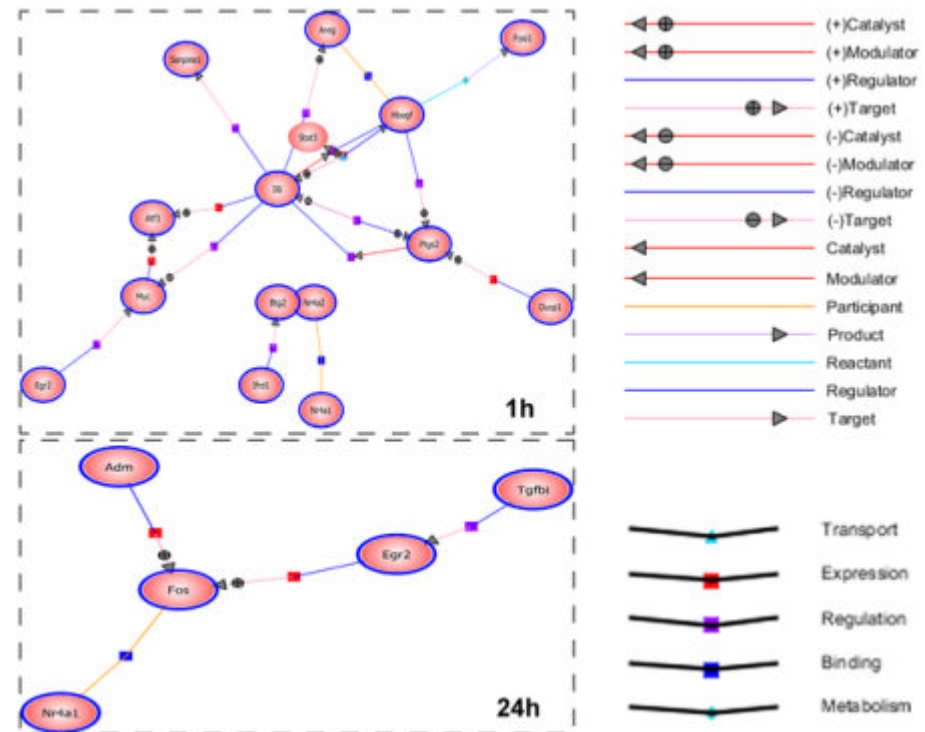


Agilent Technologies

# 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 signalling 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.



Agilent Technologies

# Combine further with

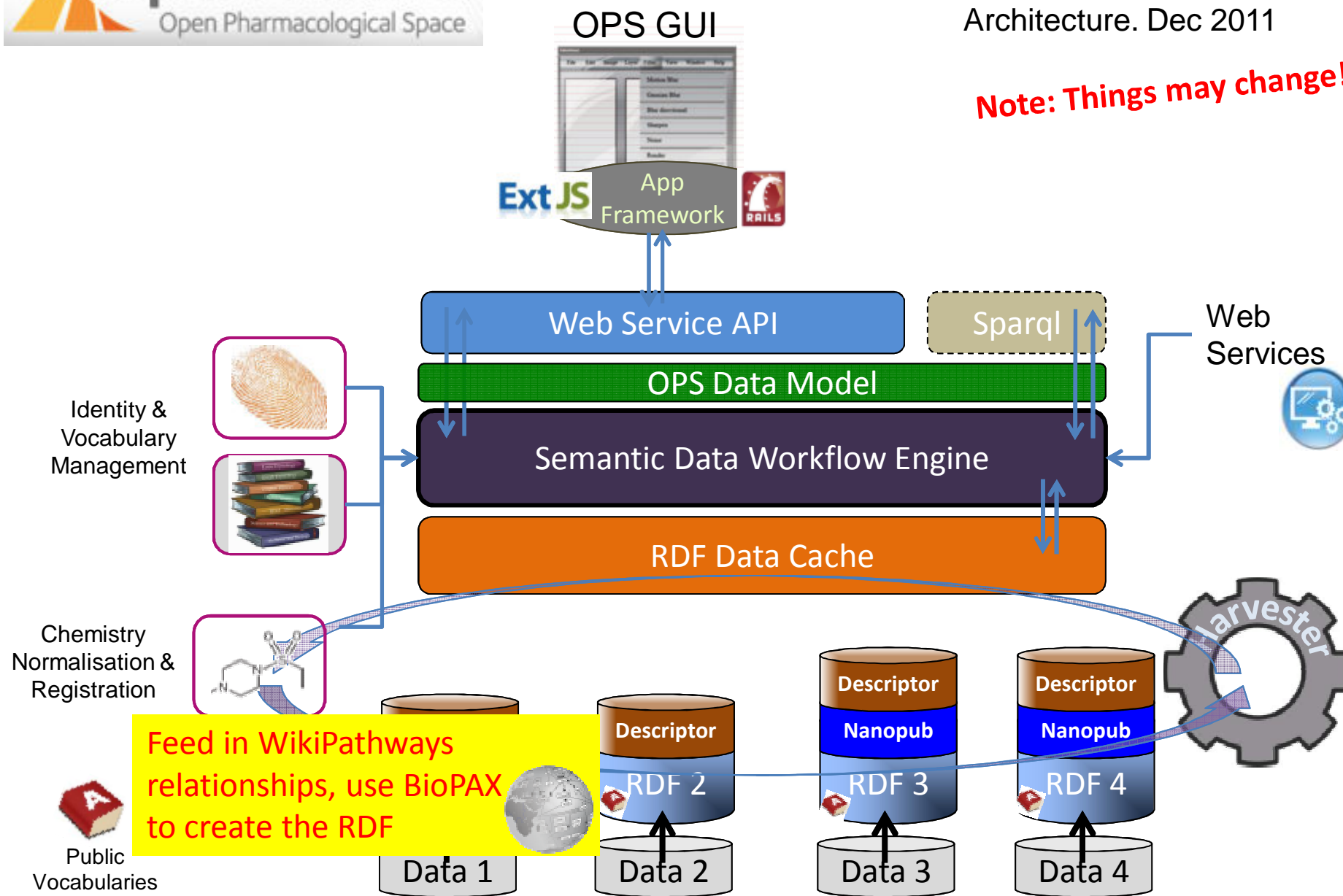
- **Open Knowledge**

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

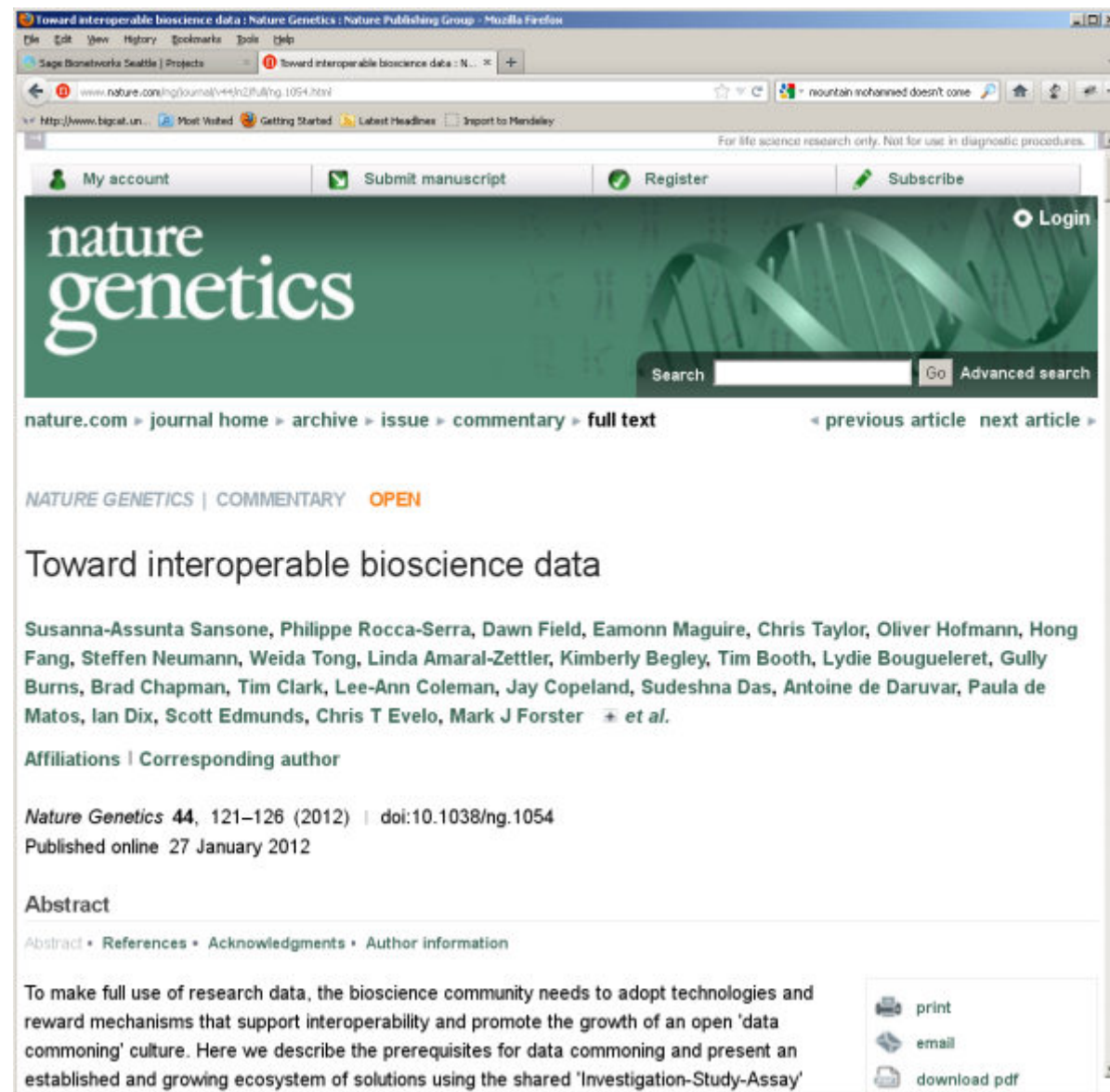
**Note: Things may change!**





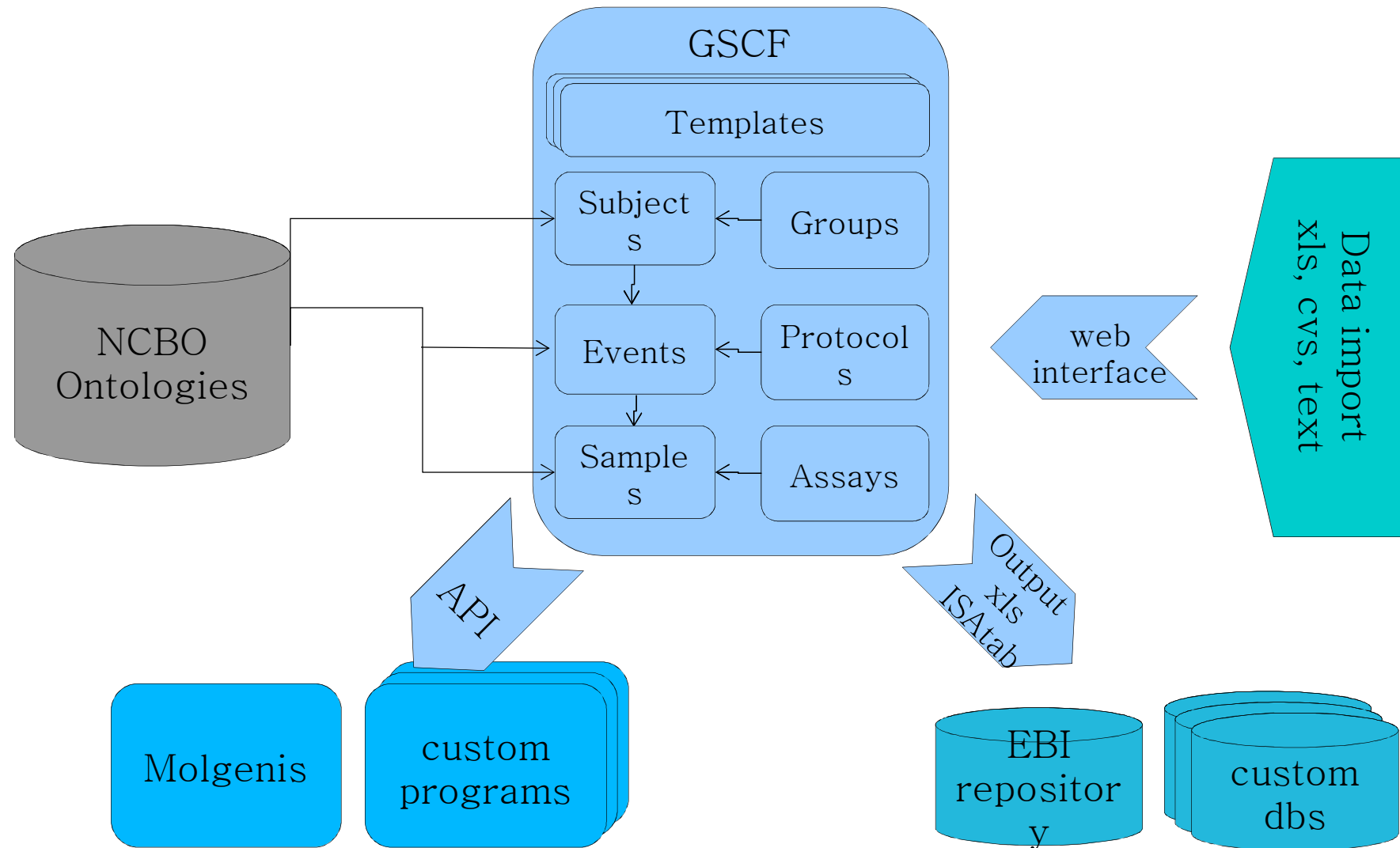
# Data capturing

Using ISA to connect to the rest of world

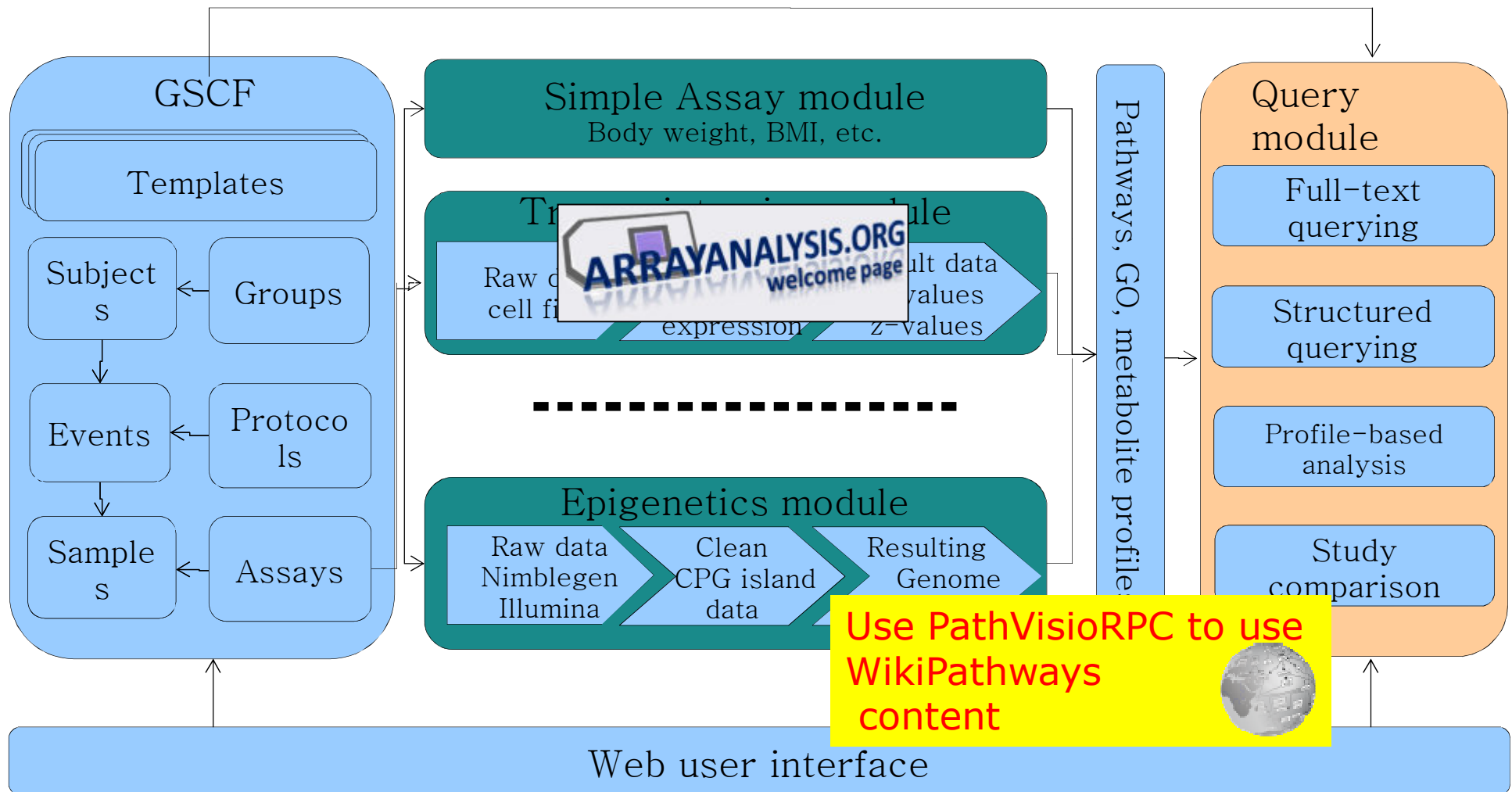


# Generic Study Capture Framework

## Data input / output



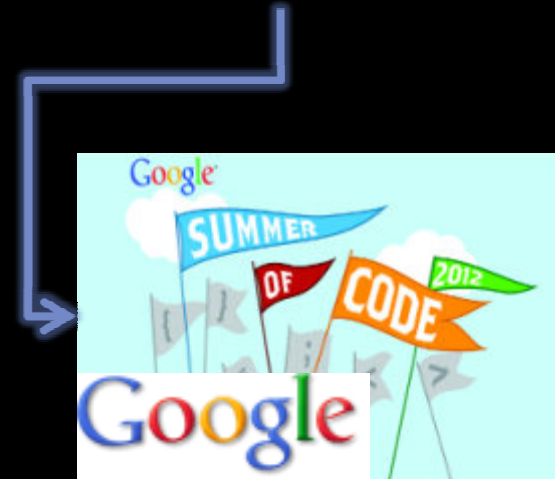
# dbNP Architecture





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

[wikipathways.org](http://wikipathways.org)



[nrnb.org](http://nrnb.org)

*Acknowledgements*